3-25-2014

Solution Nuclear Magnetic Resonance Structure and Dynamics of the Receiver Domain Sma0114

Sarah Ryan Sheftic
University of Connecticut - Storrs, sarah.sheftic@uconn.edu

Solution Nuclear Magnetic Resonance Structure and Dynamics of the Receiver Domain Sma0114

Sarah R. Sheftic

University of Connecticut, 2014

Bacterial receiver domain proteins modulate the intracellular response to external stimuli in two-component systems. There is a new subclass of receiver domains associated with Histidine Tryptophan Glutamate kinases (HWE-kinases) that have divergent sequence properties. The enzyme Sma0114 is a receiver domain that is a substrate for an HWE-kinase. The structure and dynamics of the Sma0114 protein were explored using nuclear magnetic resonance (NMR). Differences between the NMR structures of the inactive and activated states occur in helix α1, the active site loop that connects strand β3 and helix α3, and in the 455 (α4-β5-α5) face. In most receiver domains, the 455 face undergoes a structural rearrangement in the activated state to make it competent for binding downstream target molecules. Coupling between the 455 face and the active site phosphate is mediated through the rearrangement of a threonine and tyrosine residue via a mechanism called Y-T coupling. The NMR structure indicates that Sma0114 lacks Y-T coupling, and that communication between the active site and the 455 face is achieved through a conserved lysine residue that stabilizes the acyl phosphate in receiver domains. ^{15}N-NMR relaxation experiments reveal a loss of entropy due to binding ligands at the active site with compensating entropic increases in the 455 face. The dynamic character of the 455 face in Sma0114, which in part results from replacement of helix α4 by a flexible loop, may facilitate induced-fit binding of target molecules.

Solution Nuclear Magnetic Resonance Structure and Dynamics of the Receiver Domain Sma0114

Sarah Ryan Sheftic

B.S., University of Connecticut, 2009

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at the

University of Connecticut

2014

Approval Page

Doctor of Philosophy Dissertation

Solution Nuclear Magnetic Resonance Structure and Dynamics of the Receiver Domain Sma0114

Presented by

Sarah Ryan Sheftic, B.S.

Major Advisor _____

Dr. Andrei Alexandrescu

Associate Advisor _____

Dr. Debra Kendall

Associate Advisor _____

Dr. James Cole

University of Connecticut

2014

Acknowledgements

The journey of completing my dissertation and subsequent Ph.D. has not occurred in a vacuum independent of life in general. My experience has been packed with personal and academic challenges that have shaped my life inside and outside of the research arena. First, I would like to thank my mentor and advisor Dr. Andrei Alexandrescu for his invaluable help through this process. Our experiences together have been challenging and many in our department find our bickering and banter to be very entertaining. Nevertheless, we have forged a relationship based on mutual respect and a passion for research that has been the cornerstone of my career. Notably, Andrei and I enjoyed personal firsts when I was awarded the prestigious National Science Foundation graduate research fellowship. Being the recipient of such an honor has been a tremendous asset to my career in research and I could not have accomplished such a task without Andrei.

I have also received invaluable help in the form of discussions and personal recommendations from Dr. James Cole and Dr. Debra Kendall. Dr. Kendall introduced me to pre-doctoral fellowships and was the first to encourage my applications to the NSF as well as NIH. Thus Dr. Kendall became a mentor to my career early and has been an important asset on my committee who deserves more thanks and gratitude than I can express. Likewise, Dr. Cole was a patient mentor to me and helped me work through several research and grant problems in many discussions. I remember specifically that Dr. Cole took the time to comb through my reviewer comments from my initial NSF fellowship submission. For his generous donation of time and enduring my many questions I thank him whole-heartedly. Dr. Victoria Robinson has been a mentor, friend,

jokester and jack-of-all trades in my career in graduate school. Without fail Dr. Robinson has been the comedic relief in so many stressful moments. She also has edited or reviewed half of my manuscripts and provided keen insights in my projects. The last member of my committee is Dr. Daniel Gage who brought the Sma0114 project to Andrei's lab that became my main project. Without Dr. Gage I would not have three of my publications including the big-two about the structure of inactive and activated Sma0114.

Beyond the university, my research life and personal life collided in every way. As I endured the stressful life of a graduate student there were key players and events that defined my life beyond academia. First, no acknowledgements would be complete without thanking my parents Bob and Veronica who have been pillars of comfort and strength in every aspect of my life. You graciously listened to my many rants about experiments that were not working and kept me company on the phone when I would walk to my car after late-night work. Thank you for everything! To my husband Mike I owe many thanks that cannot be fully expressed. In a four-year span we put our relationship through graduate school, medical school, wedding planning, wedding execution (yes they are different), applying for new jobs and moving. I believe the top two things you should not do in the first year of marriage are change jobs and move but we decided to live life on the edge! So to my husband I thank you for sharing these tumultuous years with me and I know someday in the very distant future we will appreciate the challenges that molded us into who we are today. In the meantime, a beach vacation might be in order for relaxing and celebrating.

Table of Contents

Title page..... i

Copyright page..... ii

Approval page..... iii

Acknowledgements..... iv

List of tables..... xi

List of figures..... xii

Chapter One. Introduction..... 1

Two-component systems mediate signal transduction in bacteria..... 1

Receiver domains have a conserved structure in response regulators..... 1

Activation of receiver domains..... 4

Sma0114 deviates from known receiver domain sequences..... 5

Chapter Two. Chemical Shift Assignments for Sma0114..... 7

Abstract..... 7

Introduction..... 8

Results..... 9

Backbone chemical shift assignments..... 9

Sidechain chemical shift assignments..... 11

Secondary structure of Sma0114..... 13

Chapter Three. NMR Structure and Dynamics of Apo-Sma0114..... 15

Abstract..... 15

Introduction..... 16

Results..... 18

Solution NMR structure of apo-Sma0114..... 18

NMR relaxation data show increased flexibility for the 455 face and active site..... 25

The family of receiver domains associated with HWE-kinases is predicted to lack helix α4..... 29

Sma0114 binds Mg^{2+} and Ca^{2+}..... 29

Discussion..... 32

Chapter Four. Characterization of Activated Sma0114..... 36

Abstract..... 36

Introduction..... 37

Results..... 38

NMR structure of activated Sma0114..... 38

Conformational changes induced by the activation of Sma0114..... 42

Comparison to other receiver domains..... 44

Sidechain changes accompanying activation..... 47

Effects of Ca^{2+} and BeF_3^- binding on backbone dynamics..... 52

Discussion..... 58

Chapter Five. Materials and Methods..... 64

Sma0114 expression and purification..... 64

NMR sample preparation..... 65

Inactive Sma0114 NMR spectroscopy..... 66

Inactive Sma0114 chemical shift assignments..... 67

Chemical shifts for activated Sma0114..... 68

NMR structure determination for inactive Sma0114..... 68

NMR structure determination for BeF_3-activated Sma0114..... 69

NMR relaxation measurements..... 73

Bioinformatics analysis of Sma0114..... 74

Chapter Six. Future Directions..... 75

Appendix One. Misfolding of SEVI Inhibited by Physiological Metal Ions..... 77

Abstract..... 77

Introduction..... 78

Results..... 80

Cu^{2+} and Zn^{2+} inhibit SEVI fibrillization..... 80

Fibril morphology is retained but the amount of fibrils is lowered in the presence of Zn^{2+} and Cu^{2+}..... 84

NMR mapping identifies the two histidines in SEVI as the ligands for Zn^{2+} and Cu^{2+}..... 84

ITC shows SEVI forms oligomeric complexes with Cu^{2+} and Zn^{2+}..... 85

Discussion..... 90

Materials and Methods..... 92

Preparation of SEVI peptide..... 92

SEVI fibril formation kinetics..... 92

Transmission electron microscopy (TEM) for SEVI..... 94

NMR spectroscopy for SEVI..... 95

Isothermal titration calorimetry (ITC) for SEVI..... 96

Appendix Two. Ionization Constants of Free and Micelle-Bound Amyloid-β..... 97

Abstract..... 97

Introduction..... 98

Results..... 100

pH titration of Aβ-40 via NMR..... 100

Interpretation of pK_a values in terms of stability..... 100

Results with micelle-bound Aβ-40..... 103

Discussion..... 109

Materials and Methods..... 111

Amyloid-β preparation for NMR..... 111

pK_a measurements by NMR..... 111

Analysis of pK_a values..... 113

Appendix Three. Visualizing Amyloid-β Complexes..... 114

Abstract..... 114

Introduction..... 115

Results..... 118

Microscopy of Aβ-40 and Aβ-42 peptides..... 118

Discussion..... 125

Materials and Methods..... 127

Fluorescently tagged Amyloid-β preparation..... 127

Amyloid-β kinetics and TEM..... 127

TIRF Chamber Preparation..... 128

TIRF Microscopy..... 128

References..... 130

List of Tables

3-1 Statistics for the 20 Lowest Energy NMR Structures of Sma0114..... 19

4-1 Statistics for the 26 Lowest Energy NMR Structures of Sma0114•Ca^{2+}•BeF_3^-..... 40

A2-1 Ionization constants for Aβ-40..... 107

List of Figures

1-1 Structure of the receiver domain Spo0F..... 3

2-1 HNCACB spectra for the sequence Arg78-Glu77-Ala76-Leu75-Ile74..... 10

2-2 ^{1}H-^{15}N HSQC spectrum of Sma0114..... 12

2-3 Secondary structure of Sma0114..... 14

3-1 Fold and NMR structure of Sma0114..... 22

3-2 Active site of Sma0114..... 23

3-3 ^{15}N NMR relaxation data for Sma0114..... 27

3-4 Model-free analysis of Sma0114..... 28

3-5 Secondary structure prediction of 84 HWE-associated receiver domains..... 30

3-6 Metal-binding of Sma0114 followed by NMR spectroscopy..... 32

3-7 Perturbations of Sma0114 ^{1}H-^{15}N resonances due to metal-binding..... 33

4-1 NMR structure of the ternary Sma0114•Ca^{2+}•BeF_3^- complex..... 39

4-2 Comparison of the inactive and activated NMR structures of Sma0114..... 43

4-3 Comparison of inactive and activated states in structurally homologous single domain response regulators..... 45

4-4 Stereo-diagrams comparing selected sidechains between the inactive and activated states of Sma0114..... 48

4-5 Electrostatic potential surface for the 455 face of the inactive and activated states of Sma0114..... 51

4-6 Model-free analysis of Sma0114 dynamics based on ^{15}N relaxation data..... 54

4-7 S^2 Order parameters mapped onto the NMR structures of inactive and activated Sma0114..... 56

4-8 $R2_{ex}$ contributions derived for ternary Ca^{2+}•BeF_3^- bound Sma0114..... 57

5-1 Effects of Ca^{2+} and BeF_3^- on NMR spectra of Sma0114..... 68

A1-1 Inhibition of SEVI fibrillization..... 82

A1-2 Kinetic parameters for SEVI fibrillization..... 83

A1-3 TEM images of SEVI fibrils..... 86

A1-4 Effects of $ZnCl_2$ on the NMR spectrum of SEVI..... 87

A1-5 ITC experiments of SEVI..... 89

A2-1 ^1H-^{15}N HSQC spectrum of Aβ-40..... 101

A2-2 Titration of Glu11 using observed chemical shifts as a function of pH..... 102

A2-3 Shift in pK_a values between folded and unfolded states and the calculation of a change in folding free energy from a change in charge ($\Delta\Delta G_{titr}$)..... 104

A2-4 Titration curves of His50 from αS in various solvents..... 108

A3-1 Representative amyloid fibrillization reaction monitored by ThT fluorescence.. 117

A3-2 ThT kinetic assays for Aβ-40-488 and Aβ-42-555..... 119

A3-3 TEM images of Aβ-40-488 and Aβ-42-555..... 120

A3-4 TEM images of Aβ-40 fibrils at 11,000X magnification..... 121

A3-5 TIRF microscopy of Aβ-40-488 fibrils formed in the absence of Aβ-42-555.... 123

A3-6 TIRF microscopy of Aβ-42-555..... 124

Chapter One. Introduction

Two-component systems mediate signal transduction in bacteria

Bacterial signal transduction is commonly accomplished using two-component systems that allow bacteria to sense changes in their environment and respond using highly sophisticated phosphorylation relays (*1, 2*). The archetypal two-component system utilizes a histidine kinase that autophosphorylates itself in response to external stimuli and a response regulator that undergoes structural reorganization in response to phosphoryl transfer from its cognate histidine kinase (*1*). The histidine kinase component has two domains, an N-terminal external domain that senses environmental changes and an autokinase domain that autophosphorylates at a histidine residue. The response regulator component is located in the cytoplasm and also typically has two domains, a receiver domain that is phosphorylated by its cognate histidine kinase and an effector domain that carries out an output response that usually involves DNA binding to affect gene transcription. Receiver domains have a conserved α_5/β_5 Rossmann fold, where 5 α-helices surround 5 parallel β-sheets whereas effector domains vary among response regulators according to their downstream function.

Receiver domains have a conserved structure in response regulator proteins

Receiver domains have several conserved sequence motifs: (I) a pair of acidic residues in the β1/α1 loop ($L_{\beta1/\alpha1}$) that coordinate a divalent metal cation, (II) an aspartate at the C-terminal end of strand β3 that serves as the phosphorylation site, (III) a threonine at the C-terminal end of strand β4 that hydrogen bonds with the phosphate moiety via its sidechain hydroxyl group, (IV) an aromatic (tyrosine or phenylalanine) in strand β5 that

rotates its aromatic ring toward the hyrdophobic core upon reorientation of the threonine sidechain and (V) a lysine that flips its sidechain toward the phosphate group to provide a stabilizing salt bridge upon phosphorylation. The conserved features are highlighted on the ribbon diagram of sporulation factor Spo0F, a receiver domain from *Bacillus subtilis* in Figure 1-1. The role of the receiver domain is to accept a phosphate from its cognate histidine kinase forming a high-energy phospho-Asp bond that triggers the activation of the enzyme (*1, 3*). As such, the receiver domain acts as a "switch" that undergoes a conformational change in response to phosphorylation (*1, 2*). The effector domain typically mediates the downstream output response that is often DNA binding to affect gene expression (*1, 2*).

A smaller subset of about 15% of response regulators are single domain proteins, containing only the receiver domain (*4-6*). Of the ~30,000 non-redundant receiver domain sequences known, ~200 have been structurally characterized (*4*). Some single domain proteins housing only a receiver domain include CheY and Spo0F (*1, 7, 8*). The roles of single domain response regulators are not as well understood as their two-domain counterparts but there are three known functions. The receiver domain can act directly on a protein effector such as in the case of CheY, it can participate in a phospho-relay cascade as in the case with Spo0F, or it can act as a histidine kinase inhibitor as in the case with DivK (*5, 9, 10*). Structural studies of receiver domains have shown that these are structurally similar in their inactive apo forms but show greater variability in their phosphorylated activated states (*11*). This is likely because the function of the inactive receiver domain is to receive a phosphate from its cognate histidine kinase, whereas the activated enzyme can bind to a wide range of downstream effector proteins (*12, 13*).

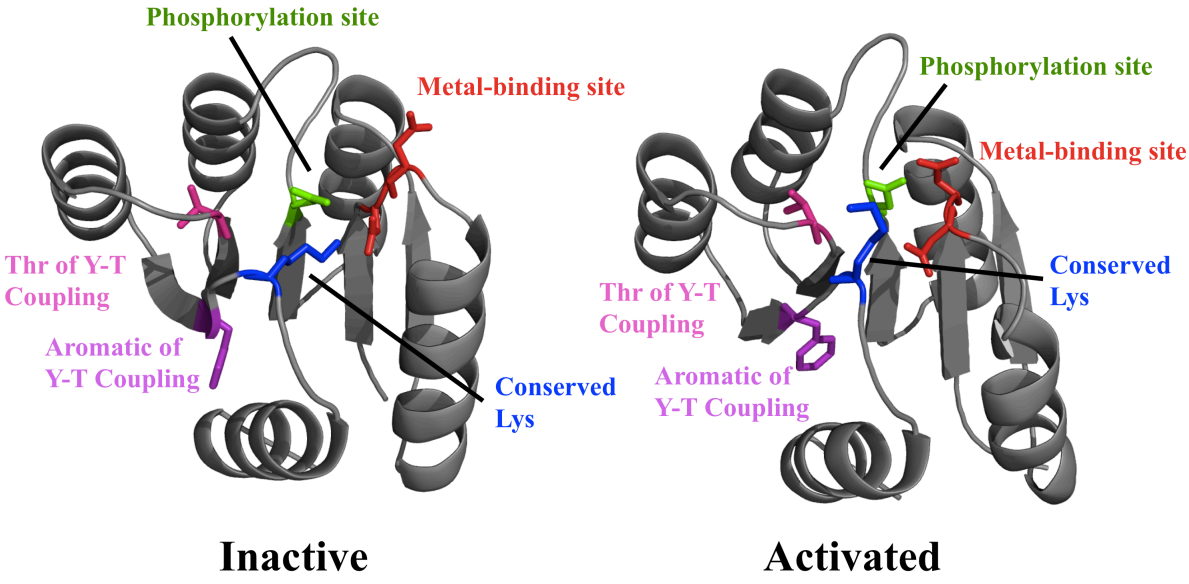

Figure 1-1 Structure of the receiver domain Spo0F. The α/β topology consists of a five-stranded parallel β-sheet surrounded by five α-helices. Conserved features of the active site are highlighed in red (I) metal binding site, in green (II) the phosphorylation site, in pink (III) the Thr of Y-T coupling, in purple (IV) the aromatic Phe of Y-T coupling and in blue (V) the Lys that forms a salt bridge with the phosphate. PDB codes 1NAT and 1PUX correspond to the inactive and activated structures respectively.

Activation of Receiver Domains

Activation of receiver domains is accomplished by phosphorylation at a conserved aspartate in the presence of a divalent metal ion such as Mg^{2+}, Mn^{2+}, or Ca^{2+} (*14*). A conserved pair of acidic residues located in the loop between strand β1 and helix α1, form part of the binding site for a divalent metal cation which is required for stabilizing the incoming phosphate group and forming the acyl-phosphate linkage (*2, 4*). The site of phosphorylation is a conserved Asp at the C-terminal end of strand β3 (*1, 2*). The C-terminal end of strand β5 houses a conserved Lys that stabilizes the incoming phosphate group (*4*). Upon phosphorylation, receiver domains undergo a conformational transition to an activated state, becoming primed for binding the downstream effector molecules that elicit the signal transduction output response. Receiver domains have autophosphatase enzymatic activity that controls the lifetime of the activated state, and hence the duration of the output response (*15, 16*).

In typical receiver domains, the phosphorylation state of the active site is communicated to the output 455 (α4-β5-α5) face of the enzyme that typically forms a binding surface for downstream targets. Activation of the protein is relayed to the 455 face through a conformational rearrangement called 'Y-T coupling', involving a structurally conserved tyrosine (Y) and threonine (T) (*1*). In this mechanism a Thr in strand β4 forms a hydrogen bond with the phosphate via its sidechain hydroxyl group. The cavity left by the Thr sidechain reorientation is filled by the reorientation of an aromatic sidechain (Tyr/Phe) from a residue in strand β5 (Figure 1-1). Upon activation the enzyme undergoes a structural rearrangement commonly in the 455 face. The receiver domain is then primed to carry out the output response.

Sma0114 deviates from known receiver domain enzymes

The two-component system encoded by the genes *sma0113* and *sma0114* from the bacterium *Sinorhizobium meliloti,* was identified in a genetic screen for altered succinate-mediated catabolite repression (*17*). *S. meliloti* is a gram-negative soil bacterium in the α-proteobacteria. The bacterium can exist in the soil as a free-living organism, or it can live as a nitrogen-fixing symbiont in specialized organs called nodules, that it induces on the roots of certain legumes such as alfalfa. Based on previous work it is expected that *sma0113* and *sma0114* play a role in *S. meliloti* catabolite repression and polyhydroxy butyrate synthesis (*17*). The tandem genes encode a histidine-tryptophan-glutamate (HWE) histidine kinase (Sma0113), and a single-domain response regulator (Sma0114).

The architecture of the classic histidine kinase involves two modular domains termed the dimerization and catalytic domains. The dimerization domain is a helix bundle and the catalytic domain has a mixed α/β topology with five β-strands and three α-helices. In archetypal histidine kinases the H box that houses the conserved His for autophosphorylation is located in the dimerization domain (*1*). The catalytic domain is responsible for ATP binding and contains several conserved sequence motifs. The N, G1, G2 and F boxes are amino acid motifs involved in nucleotide binding (*1, 2*).

HWE-kinases were first described in 2004 (*18*) and constitute a subclass of the larger histidine kinase superfamily. Compared to the well-characterized histidine kinase family, HWE-kinases have an altered ATP binding site, which lacks the F box that is normally an integral component of the ATP lid. There is a His in what appears to be the H box of the HWE-kinase, however, there is little alignment between this H box and

known H box sequences. Furthermore, there is a His in the N box as well as Trp and Glu residues in the G1 box of the protein (*18*). The His, Trp and Glu resides in the H and G1 boxes are essential for autophosphorylation of the HWE-kinase family.

Preliminary sequence analysis of Sma0114 reveals that it adopts a prototypical Rossmann-fold but has features that set it apart from other receiver domains. Sma0114 has many conserved residues typically present in receiver domains including the Asp (Asp60) that accepts a phosphate from its cognate HWE-kinase Sma0113 and a pair of acidic residues involved in prototypical metal binding (Glu15 and Asp16). One interesting feature of Sma0114 is the presence of a new motif PFxFATGY located in the beta strand that houses the Thr typically involved in the classic Y-T coupling mechanism. Although the conserved Thr is present in the sequence, Sma0114 does not have the conserved aromatic residue that is the other half of this coupling motif.

The NMR investigations described herein show that the unusual sequence features of Sma0114 lead to unique structural features, primarily affecting the 455 face of the enzyme. ^{15}N relaxation data show that there are also changes in the dynamics of the 455 face of Sma0114 compared to the corresponding regions of other receiver domains (*19, 20*). By contrast the active site of the enzyme and the ability to bind metals is retained, suggesting that the unusual features of Sma0114 do not alter the activation mechanism but rather the conformational changes of the 455 face accompanying phosphorylation.

Chapter 2. Chemical Shift Assignments for Sma0114

Abstract

Response regulators of bacterial two-component systems undergo extensive structural reorganization in response to phosphoryl transfer from their cognate histidine kinases. The response regulator encoded by the gene *sma0114* of *S. meliloti* is a part of a unique class of two-component systems that employ HWE histidine kinases. The distinct features of Sma0114 include a PFxFATGY motif that houses the conserved threonine in the "Y-T coupling" conformational switch and the replacement of the conserved Phe/Tyr in Y-T coupling by a Leu. Presented here are ^{1}H, ^{15}N, and ^{13}C NMR assignments for Sma0114. Chemical shift assignments are a pre-requisite for an NMR study. Secondary structural elements of the protein were identified based on TALOS chemical shift analysis, $^{3}J_{HNHa}$ coupling constants and hydrogen-deuterium exchange. The secondary structure determined by NMR is in good agreement with that predicted from the sequence.

Introduction

Receiver domains of response regulators are typically about 110 amino acids long and are structurally and functionally analogous to members of the Ras family of small eukaryotic GTPases (*21*). It was expected that the 120 amino acid (13.5 kDa) Sma0114 protein would have a CheY-like structure, belonging to the α/β Rossmann-fold superfamily. Regulatory regions of Sma0114 were identified via comparison to known receiver domains and include a set of conserved residues clustered in an active site at the C-terminal edge of the β-sheet. Two additional conserved amino acids include a hydroxylic residue (Ser/Thr) and an aromatic residue (Phe/Tyr) that play direct roles in communication between the active site and output response α4-β5-α5 (455) face of the protein. The Y-T coupling mechanism thus transforms the overall character of the 455 face of the protein. The conformational transition in turn governs the specific interactions that mediate the output response (*22-24*).

Sma0114 is a member of a recently identified class of response regulators associated with HWE-kinases. Response regulators associated with the HWE kinases (*25*) either because the two components are linked on the same polypeptide chain or encoded in the same operon like the Sma0113(histidine kinase)/114 (response regulator)pair, often have a conserved motif at the C-terminal end of strand β4. Response regulators containing this PFxFATGY, or "FAT GUY" motif have not been structurally characterized. To better understand the unique features of Sma0114 including the non-conserved Y-T coupling mechanism, the role of the PFxFATGY motif, and activation by its cognate HWE kinase an NMR characterization of the structure and dynamics of the protein was undertaken.

Results

Backbone Chemical Shift Assignments

A series of NMR experiments were used to obtain chemical shift assignments for the enzyme. Several through-bond (TOCSY-type) experiments were used initially to establish backbone resonance assignments in Sma0114. Specifically, two redundant sequential walk pathways were used. Using multiple pathways ensured the accuracy of the resonance assignments. The first pathway used Cα connections via HNCA and HN(CO)CA experiments. The second pathway used backbone carbonyl connections via HNCO and HN(CA)CO experiments. HNCACB experiments were used in conjunction with the Cα and carbonyl sequential walk methods. The HNCACB experiment was used to correlate the Cα and Cβ resonances of a single residue using through-bond coupling to the Cα and Cβ of the preceding residue in the sequence. Using this method, full chemical shift assignments were obtained. Figure 2-1 illustrates example resonances from the HNCACB spectrum of the inactive enzyme. Using these multidimensional NMR experiments backbone amide resonances were assigned in a 2D ^{1}H-^{15}N HSQC of Sma0114. The 2D ^{1}H-^{15}N HSQC spectrum is unique to a polypeptide chain and provided a platform with which to characterize other atomic assignments. The ^{1}H-^{15}N HSQC experiment is the most ubiquitously used spectrum to identify a protein and is often used to establish amino acids involved in substrate binding. Backbone ^{1}H-^{15}N assignments for apo-Sma0114 are summarized in Figure 2-2.

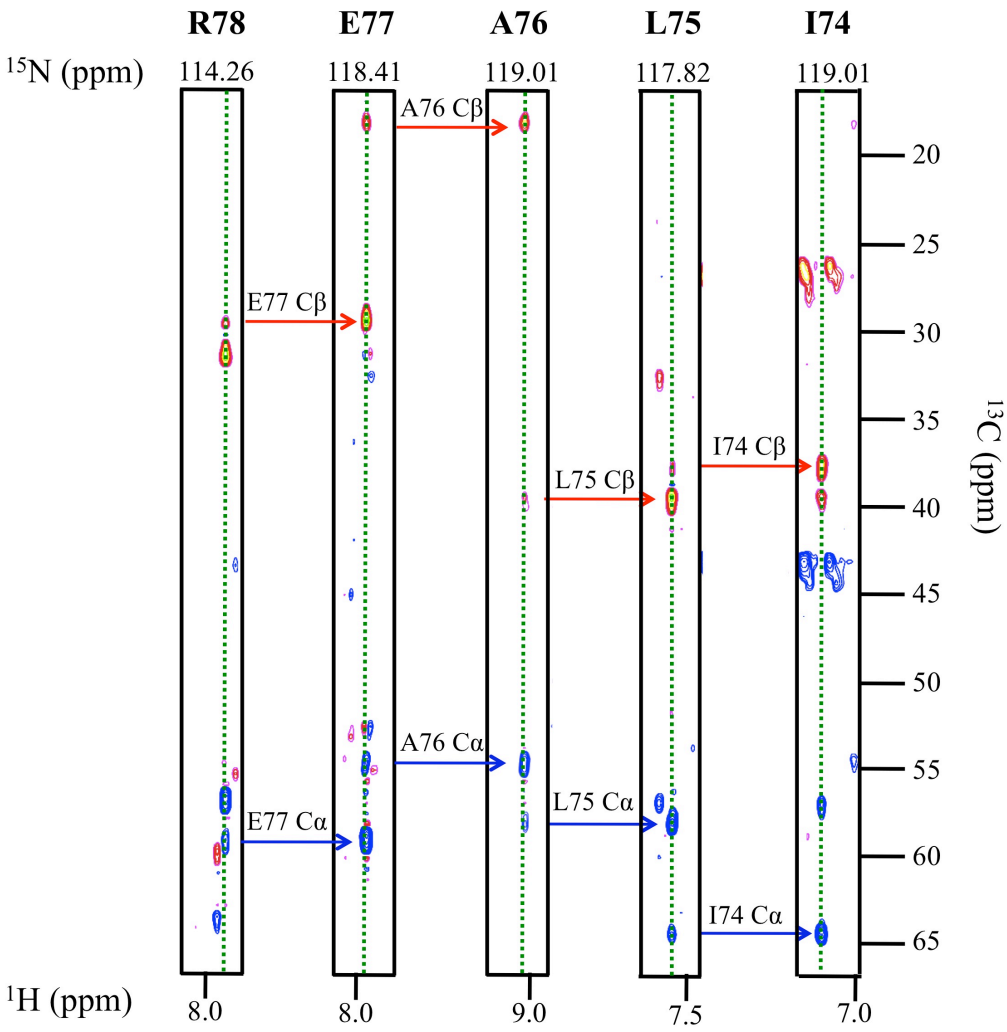

Figure 2-1 HNCACB Spectra for the sequence Arg78-Glu77-Ala76-Leu75-Ile74. Greater intensity Cα and Cβ peaks in each panel correspond to the residue X. Lower intensity Cα and Cβ resonances belong to the residue X-1 in the amino acid sequence. For example, the R78 panel contains contours corresponding to the Cα and Cβ chemical shifts for Arg78 as well as lower intensity Cα and Cβ resonances belonging to Glu77. In this sequential walk manner ^{1}H and ^{15}N as well as Cα and Cβ chemical shifts were determined. Dotted green lines in each panel identify the spin systems. Blue and red resonances represent Cα and Cβ peaks respectively.

Sidechain Chemical Shift Assignments

Sidechain resonances were also assigned using 3D through-bond correlation (TOCSY) experiments. Manual assignments using HCCH-TOCSY experiments, described by Kay, L.E. et al. were employed first, however, the lower resolution of the experiments made unambiguous assignments difficult and tedious. For example, resolving methyl group ^{1}H resonances to unique residues was particularly challenging. In order to make unambiguous assignments for aliphatic sidechain ^{13}C and ^{1}H resonances we used the approach described by Sun, Z.Y. et al (*26*). The described 3D HCCONH and CCONH experiments correlate aliphatic proton and carbon resonances with the HN group of each residue in the sequence. The HN correlation provides the unambiguous avenue for sidechain assignments as each amino acid has one unique HN chemical shift (Figure 2-2). Sidechain chemical shifts were necessary to provide comprehensive assignments for the enzyme to be used in through-space or NOESY-type experiments. Unambiguous assignments helped improve the precision of the NMR structure of the enzyme.

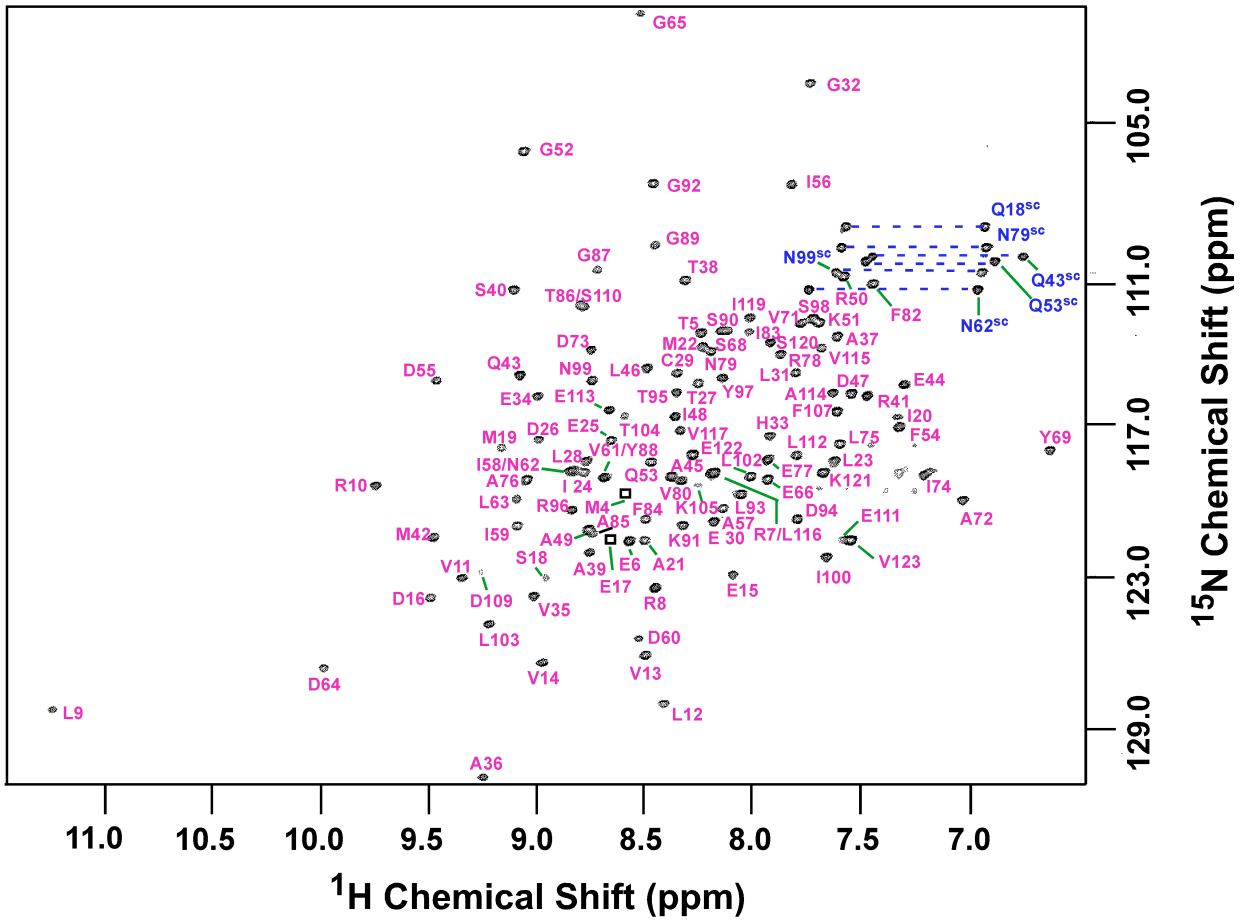

Figure 2-2 ^{1}H-^{15}N HSQC spectrum of Sma0114. Backbone assignments are indicated with pink labels. Peaks that could only be seen at contours lower than shown (M4, E17) are denoted by squares. Side chains amide resonances are indicated with blue dashed lines and labels, with the superscript SC. Data were collected at pH 6.0 and a temperature of 37°C.

Secondary Structure of Sma0114

Following chemical shift assignments for backbone and sidechain residues, secondary structure predictions were performed using PSIPRED (*27*). These data are summarized in Figure 2-3. The secondary structure established by NMR agrees with the secondary structure predicted from the sequence using the PSIPRED server (*27*). Both methods indicated that helix α4 of the prototypical CheY-fold, was absent in Sma0114 (corresponding to residues 90-99). The 455 face of response regulators is critical in transmitting the kinase phosphorylation signal to downstream binding partners (*24, 28*), so the absence of helix α4 in Sma114 led to the conclusion that there was a new distinct mechanism of coupling between HWE kinases and their cognate response regulators.

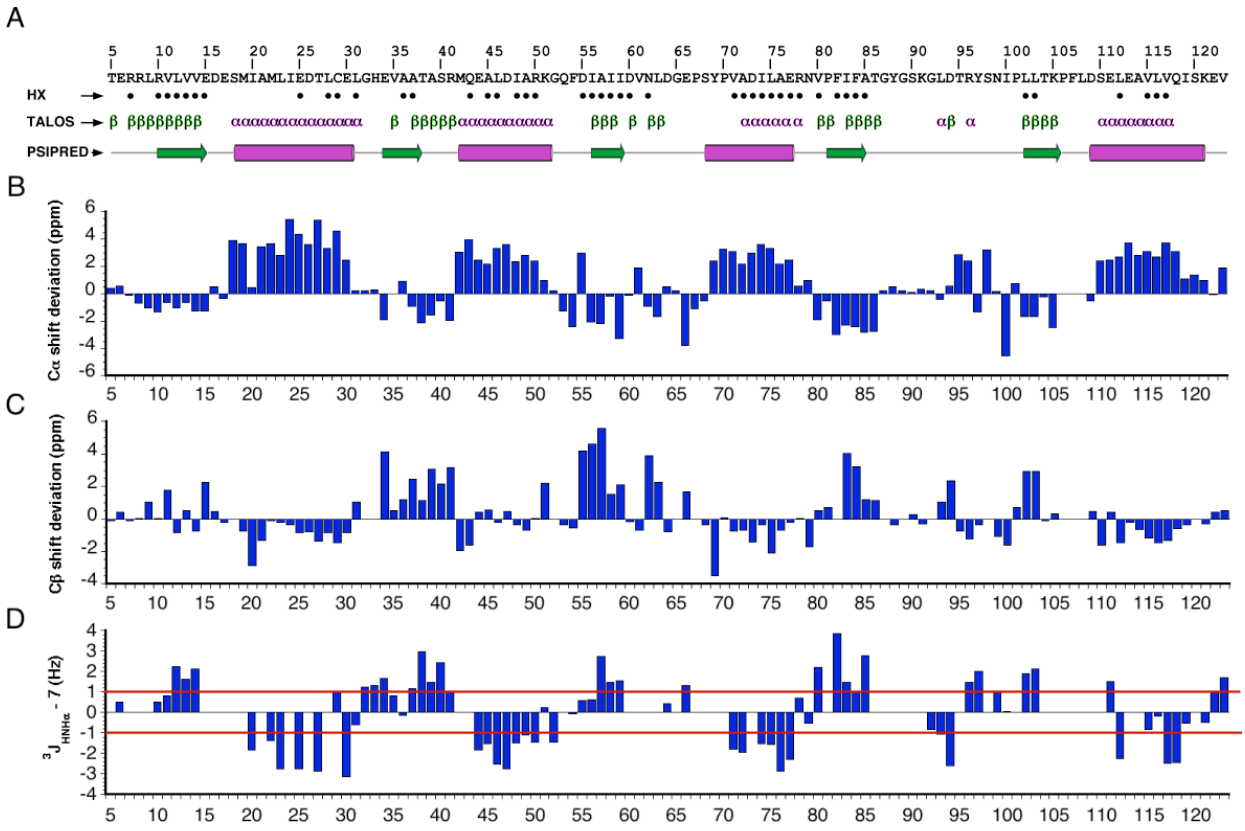

Figure 2-3 Secondary structure of Sma0114. (**A**) Amino acid sequence of Sma0114, amide protons protected from exchange after 2 hour in D_2O at pH 6.0 and 37 ^oC, TALOS predictions of secondary structure based on HN, N, Hα, Cα, Cβ and C' chemical shifts (*29*), and secondary structure predicted using the PSIPRED server (*27*). (**B**) Cα chemical shifts minus random coil values (*30*). Positive deviations correspond to α-helix, negative to β-sheet. (**C**) Cβ chemical shifts minus random coil values (*30*). Negative deviations correspond to α-helix, positive to β-sheet. (**D**) $^3J_{HNH\alpha}$ couplings obtained from a 3D HNHA experiment, presented as $^3J_{HNH\alpha}$ – 7 Hz. Positive and negative differences correspond to β-sheet and α-helix secondary structure, respectively.

Reprinted from Sheftic et al. 2012 (*31*)
I performed all experiments except the secondary structure prediction Figure 3-5

Chapter Three. NMR Structure and Dynamics of Apo-Sma0114

Abstract

The solution NMR structure and dynamics of Sma0114 from the bacterium *S. meliloti* is described. This is the first such characterization of a receiver domain from the HWE-kinase family of two-component systems. The structure of Sma0114 adopts a Rossmann-fold but has features that set it apart from other receiver domains. Sma0114 lacks the fourth α-helix of the consensus 455 face and ^{15}N relaxation data show that it is replaced by a segment that is flexible on the ps-ns timescale. Secondary structure prediction of Sma0114 and other HWE-kinase-associated receiver domains suggests that the absence of helix α4 may be a conserved property of this family. In spite of these differences, Sma0114 has a conserved active site, binds divalent metal ions such as Mg^{2+} and Ca^{2+} that are required for phosphorylation, and exhibits μs-ms active site dynamics similar to other receiver domains.

Introduction

In order to correlate the new sequence features of Sma0114 with changes in tertiary structure we solved the solution NMR structure of the enzyme. Presented here is the 3-dimensional structure of Sma0114 in the apo state. The NMR data support the secondary structure predictions presented in Chapter 2 and prove that the topology of the protein deviates from an α_5/β_5 fold to an α_4/β_5 fold. The missing α-helix maps to the area of the protein that based on sequence alignments would house helix α4, the first constituent of the 455 face.

^{15}N T1, T2, and NOE relaxation data were used as input for a Model-Free simulation that provides information on the amplitude (S^2 order parameter) and timescale of bond motions (*32*). Dynamics data are compared between the inactive and activated protein (Chapter 4) to assess the role of flexibility in enzyme function. Dynamics of the apo-protein reveal that the loop replacing helix α4 between strands β4 and β5 is flexible on the ps-ns timescale, indicative of fast motions similar to those seen for protein N and C-termini. Analysis of μs-ms dynamics reveal four distinct regions of the protein that have increased $R2_{ex}$ terms, indicative of conformational exchange.

We performed a metal titration in order to determine saturating concentrations of a divalent cation required to stabilize the acyl-phosphate linkage in the active site of Sma0114. Our initial profile of chemical shift changes in ^{1}H-^{15}N HSQC spectra upon Mg^{2+} binding showed a weak affinity for the metal ion. The unusually high (150 mM) concentration of $MgCl_2$ required to saturate the metal binding site led to a loss in signal/noise during data acquisition presumably due to the increase in salt concentration within the sample. In order to circumvent the use of high metal concentrations, $CaCl_2$ was

titrated into Sma0114 samples in order to determine the affinity of the enzyme for Ca^{2+}. A concentration of 1.5 mM $CaCl_2$ was required to saturate the metal binding site indicating preferential binding of Ca^{2+} for Sma0114.

The NMR investigations described herein show that the unusual sequence features of Sma0114 lead to unusual structural features, primarily affecting the 455 face of the enzyme. ^{15}N relaxation data show that there are also changes in the dynamics of the 455 face of Sma0114 compared to the corresponding regions of other receiver domains (*19, 20*). By contrast the active site of the enzyme and the ability to bind metals is retained.

Results

Solution NMR Structure of Apo-Sma0114

Structure statistics for the NMR models calculated are given in Table 3-1. Figure 3-1 shows the NMR structure of Sma0114. The overall fold of the protein is similar to the canonical α_5/β_5 Rossman-fold of other receiver domains, which is schematically illustrated in Figure 3-1A. Backbone (C^{α}, N, C') traces of the 20 lowest energy NMR structures of Sma0114 are shown in Figure 3-1B.

The NMR structure of Sma0114 closest to the ensemble average is compared with that of a typical receiver domain Spo0F in Figures 2C and 2D, respectively. The deviations from the canonical receiver domain fold occur primarily along the 455 face of Sma0114. The most conspicuous difference in Sma0114 (Figure 3-1C), is that helix α4 is replaced by a disordered segment (residues 88-93). Another difference, is that strand β5 which is shorter than the other β-strands due to two flanking proline residues (P101, P106), turns inward toward the core of the Sma0114 structure (Figure 3-1C).

To look more closely for similarities and differences to other receiver domains, the DALI server (*33*) was used to find the 10 best structural matches to Sma0114. The 10 best matches were all structures of inactive states of receiver domains, and had an average backbone RMSD of 3.17 Å to our Sma0114 structure over an average alignment length of 99 residues. The most prominent difference in Sma0114 is the absence of helix α4 (Figure 3-1C). Another important change is that the loop following strand β3 appears to cover and possibly restrict access to the phosphorylation site residue, Asp60 (Figure 3-1C), whereas it extends away from the hydrophobic core in the homologous structures (Figure 3-1D).

Table 3-1: Statistics for the 20 Lowest Energy NMR Structures of Sma0114

NMR restraints (total)	1627	
Distance (total)	1431	
Intraresidue NOEs	422	
Sequential NOEs	486	
Short range NOEs (1< |i-j| < 5)	109	
Long range NOEs (5 < |i-j|)	414	
Hydrogen bonds (32x2)	64	
Dihedral (ϕ 61, ψ 54, χ_1 17)	132	
Residual restraint violations		
NOE (Å)[1]	0.039 ± 0.004	
Dihedral (°)[2]	0.539 ± 0.018	
RMS deviations from ideal geometry		
Bonds (Å)	0.0040 ± 0.0003	
Angles (°)	0.73 ± 0.02	
Improper torsions (°)	0.54 ± 0.02	
Van der Waals energy (kcal/mol)[3]	86.28 ± 2.37	
Lennard-Jones energy (kcal/mol)[4]	-83.18 ± 7.40	
ProcheckNMR Z-score[5]	-2.52	
Ramachandran plot		
most favored (%)	82.3	
allowed (%)	15.3	
generously allowed (%)	1.8	
disallowed (%)	0.6	
Coordinate RMS deviations (Å)		
NMR ensemble to average	C_α, C, N	All Heavy
Entire domain (110 residues)[6]	0.94	1.46
Excluding loops (72 residues)[7]	0.86	1.38

[1] Structure contains no NOE violations greater than 0.3Å
[2] Structure contains no dihedral violations greater than 5°
[3] E_{vdw} was calculated using the X-PLOR *F*repel function (*34*) with van der Waals interactions and atomic radii set to 0.8 times their CHARMM (*35*) values.
[4] *E*L-J was calculated using the CHARMM empirical energy function (*35*)
[5]ProcheckNMR used via Protein Structure Validation Suite (PSVS) (*36*)
[6] Excluding N-terminus (residues 1-9) and C-terminus (residues 118-123)
[7] Only residues in regular secondary structure (10-15, 19-31, 34-39, 42-51, 56-60, 70-78, 81-86, 102-105, 108-117)

The alignment to structurally homologous proteins using the DALI server (*33*) also indicates that the C-terminal portion of the PFxFATGY motif in Sma0114 differs from the corresponding region in the canonical receiver domain fold. The PFxFATGY motif is highlighted in cyan on the NMR structure of Sma0114 in Figure 3-2A. This motif encompasses all of strand β4 and forms the beginning of the loop between strands β4 and β5. The conserved Thr86 of the Y-T coupling pathway is part of this motif and is the last residue in strand β4 (shown in purple in Figure 3-2A). The partner aromatic residue in the Y-T coupling mechanism is replaced by a Leu103 in strand β5 (shown in black in Figure 3-2A). The absence of the conserved aromatic residue has also been seen in the single domain response regulator CheY2 from *S. meliloti,* an enzyme that has been reported to lack Y-T coupling (*37*). These changes provide compelling evidence that the prototypical Y-T coupling mechanism is altered in Sma0114.

The two residues of the PFxFATGY motif that form part of the loop immediately following strand β4 (Gly87 and Tyr88) show large structural differences in Sma0114 compared to the top 10 structural similarity hits. These two residues are located immediately before the start of the disordered segment that replaces helix α4 in Sma0114. The C-terminal portion of the PFxFATGY motif in Sma0114 superimposes with the N-terminal portion of helix α4 in the 10 closest structural matches. Together with the placement of the Y-T coupling threonine, this suggests that at least the C-terminal part of the PFxFATGY motif belongs to the 455 face of Sma0114.

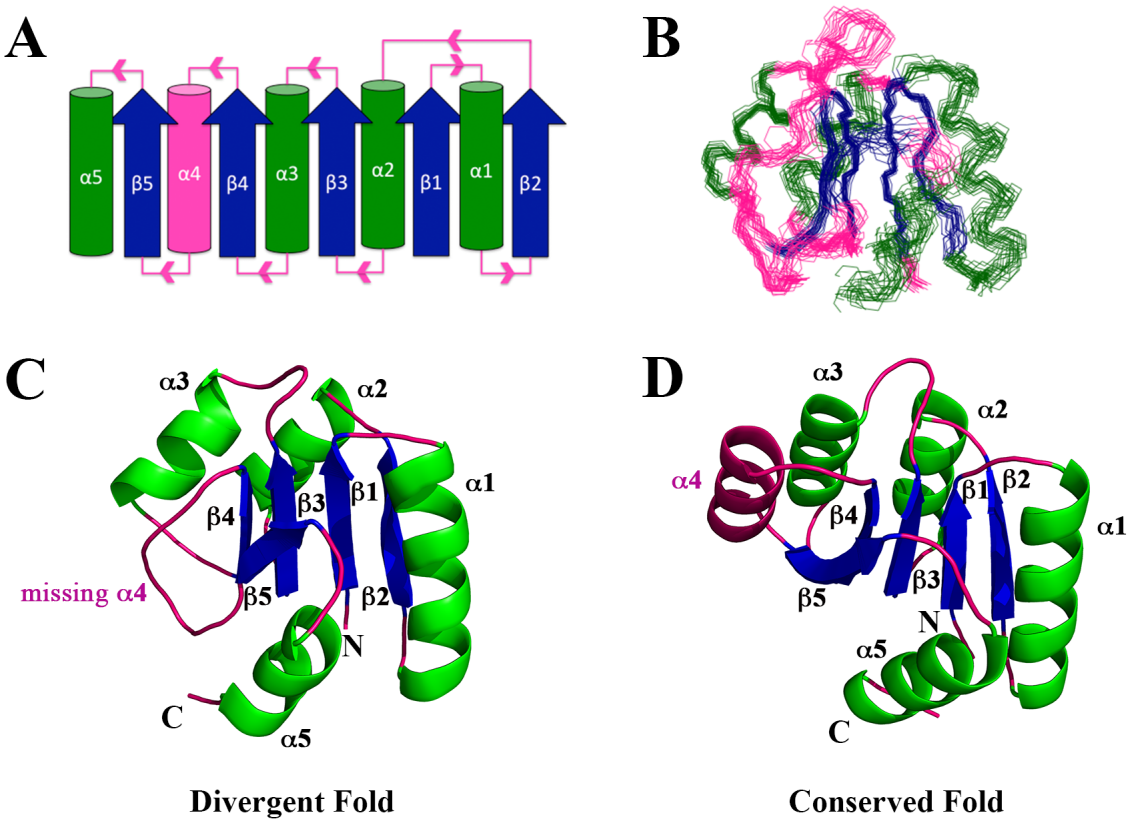

Figure 3-1 Fold and NMR structure of Sma0114. (**A**) Topology diagram for the canonical α_5/β_5 Rossman fold of receiver domains. Parallel β-sheets (blue) are surrounded by α-helices (green) with pink segments indicating loops. The fourth α-helix which is disordered in Sma0114 is also indicated in pink. (**B**) Backbone representation of the 20 lowest energy NMR structures. Regular secondary structure elements 9-15 (β1), 19-31 (α1), 34-39 (β2), 42-51 (α2), 55-60 (β3), 70-78 (α3), 81-86 (β4), 102-105 (β5), 110-117 (α5) were used to superimpose the structures. (**C**) NMR structure of Sma0114 closest to the ensemble average. The fourth α-helix is replaced by a disordered segment. (**D**) NMR structure of Spo0F (PDB accession code 1NAT) (*38*). The fourth α-helix is shown in pink. N and C termini are noted in (**C**) and (**D**) with N and C respectively.

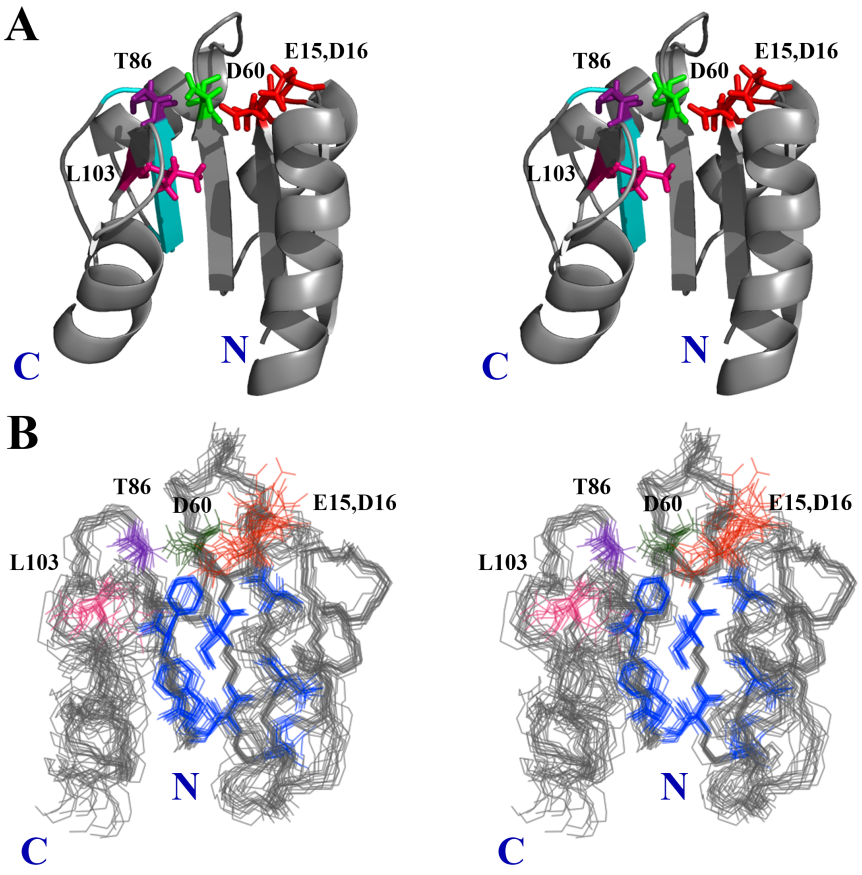

Figure 3-2 Active site of Sma0114. (**A**) Stereo-diagram of the NMR structure of Sma0114 closest to the ensemble average, showing selected sidechains of active site residues: metal-ligands Glu15 and Asp16 in red, the phosphorylation site Asp60 in green, and Thr86 and Leu103 of the Y-T coupling pair in purple and pink, respectively. The PFxFATGY motif is shown in cyan. (**B**) Stereo-view of the 20 lowest energy NMR structures of Sma0114 illustrating the precision of sidechains from different parts of the molecule. Sidechains comprising the hydrophobic core (blue) are well defined while those of active-site residues (color scheme as above) show poorer precision due to increased dynamics on the µs-ms timescale.

While the 455 face of the enzyme shows deviations from the receiver domain superfamily, the active site of Sma0114 has features consistent with other receiver domains (Figure 3-2A). The metal binding residues, Glu15 and Asp16, are located at a conserved position in the loop between β1 and α1 (*11, 13, 39*). The phosphorylation site (Asp60) is at the C-terminal end of β3 like in other receiver domains. The hydroxylic residue (Thr86) that hydrogen bonds with the incoming phosphate in the Y-T coupling mechanism is located at the C-terminal end of strand β4. The basic residue Lys105, which typically stabilizes the activated state by forming a salt bridge with the incoming phosphate occurs in a conserved location in strand β5.

Strands β1-β4 and helices α1-α3 have the highest precision in the NMR structure (Figure 3-2B). Regions with lower precision occur in the 455 face and include the segments corresponding to helix α4, strand β5, and helix α5. The active site of Sma0114 also has lower precision in the NMR structure, as illustrated in Figure 3-2B which compares sidechains of residues in the active site and in the hydrophobic core of the protein. The equilibrium shift theory of activation hypothesizes that receiver domains exist in their inactive and active substates simultaneously in solution. Activation via phosphorylation shifts the population from the inactive to the active form (*4, 20*). It follows that that the active site, which is experiencing exchange between inactive and active conformational substates on a μs-ms timescale, would have a lower precision due to $R2_{ex}$ line broadening. To verify that the lower precision of the active site and 455 face in the NMR structure of Sma0114 is due to genuine flexibility, as it is in other receiver domains (*39, 40*), we characterized the backbone dynamics of the protein using ^{15}N relaxation data.

NMR Relaxation Data Show Increased Flexibility for the 455 Face and Active Site

Relaxation data for Sma0114 are shown in Figure 3-3. The R1 values are roughly constant over the length of the protein with a mean value of 2.00 ± 0.01 s^{-1}, except for raised rates in the region between residues 89 and 99, which also shows lowered ^{1}H-^{15}N NOEs (Figure 3-3A). The R2 data reveal four distinct regions that have relaxation contributions greater than the mean value of 8.0 ± 0.3 s^{-1}, as indicated in Figure 5B. The mean ^{1}H-^{15}N NOE value is 0.75 ± 0.01, close to the theoretical maximum of 0.80 (*41*). Lowered ^{1}H-^{15}N NOEs are seen for the chain termini and the region between residues 88 and 95 that corresponds to the missing helix α4 (Figure 3-3C).

To interpret the relaxation data in terms of backbone dynamics we performed a Model-Free analysis (*42*) using the program *Tensor2* (Figure 3-4). We obtained a correlation time of 5.1 ns for Sma0114, consistent with a monomeric protein of 13.5 KDa. The S^2 values which describe the amplitudes of internal motions on the ps-ns timescale are shown in Figure 3-4A and are mapped on the NMR structure in Figure 3-4B. Except for the chain termini, the only region in the protein with low S^2 order parameters is the flexible segment that replaces helix α4 (Figure 3-4A). The $R2_{ex}$ terms that describe exchange contributions to R2 relaxation from dynamics on the μs-ms timescale are shown in Figure 3-4C and are mapped on the NMR structure in Figure 3-4D. Residues that experience significant $R2_{ex}$ contributions cluster to four region of the protein: (I) the metal binding site, (II) the phosphorylation site, (III) Thr86 which is predicted to hydrogen bond with an incoming phosphate in the Y-T coupling mechanism, and (IV) Lys105 which stabilizes the incoming phosphate through a salt-bridge. Thus all four regions with $R2_{ex}$ contributions are in the active site of the enzyme, which has been

shown for other receiver domains to be subject to μs-ms timescale interconversion between inactive and active substates as described in the equilibrium shift theory of activation (*19, 20*).

Despite their strong structural homology, receiver domains from various two-component systems exhibit different dynamics on the μs-ms timescale. $R2_{ex}$ terms for Spo0F, an intermediate in the sporulation relay of *B. subtilis*, are larger for the metal binding site compared to the 455 face of the enzyme (*19*). Conversely, in NtrC, the nitrogen regulatory response regulator from *S. typhimurium*, $R2_{ex}$ contributions are large for the 455 face of the enzyme whereas the metal binding site shows no exchange broadening (*6, 20*). The $R2_{ex}$ profile of Sma0114 differs from Spo0F and NtrC in that it shows exchange broadening not only in the metal-binding and phosphorylation sites but also in the β5 and α5 regions of the conserved 455 face.

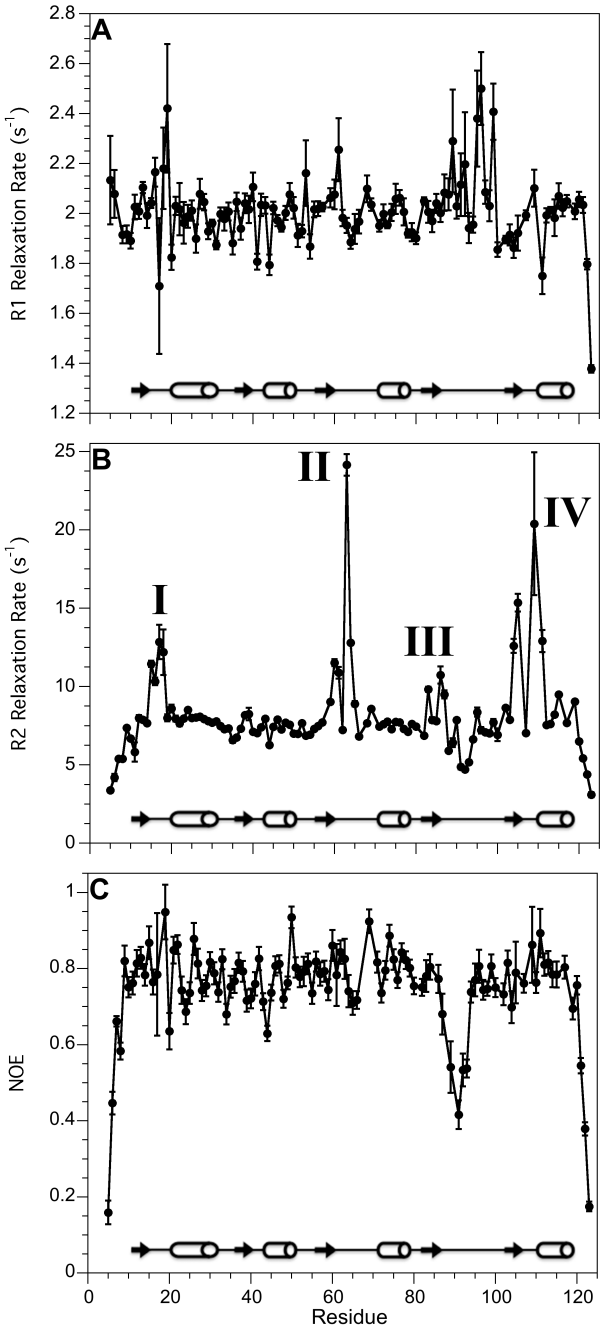

Figure 3-3 ^{15}N NMR relaxation data for Sma0114. (**A**) R1 rates. (**B**) R2 rates. (**C**) ^{1}H-^{15}N NOE values. Error bars are shown for all data points but in some cases are smaller than the symbols used to depict the data. The secondary structure of Sma0114 is indicated with arrows (β-sheets) and cylinders (α-helices) in each panel.

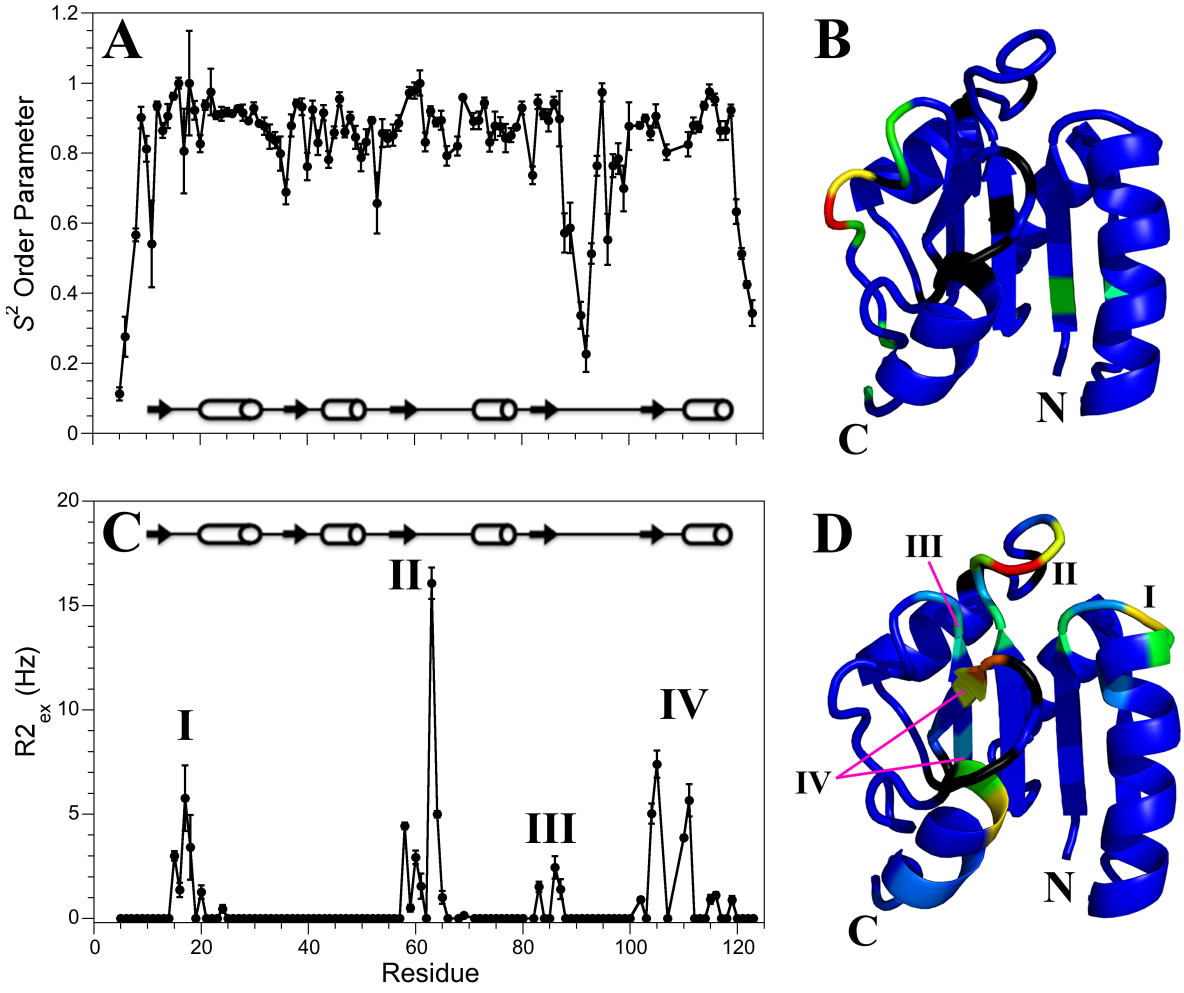

Figure 3-4 Model-free analysis of Sma0114. (**A**) S^2 order parameters. (**B**) S^2 order parameters mapped onto the NMR structure of Sma0114. The color map goes from blue (rigid) to red (flexible). (**C**) $R2_{ex}$ values. (**D**) $R2_{ex}$ values mapped onto the NMR structure of Sma0114 with colors ranging from blue (negligible $R2_{ex}$ contributions) to red (large $R2_{ex}$ contributions). Residues for which no data were available are shown in black.

The Family of Receiver Domains Associated with HWE-Kinases is Predicted to Lack Helix α4

The NMR structure of Sma0114 is the first of an HWE-kinase associated receiver domain. Due to the unusual features of this enzyme we performed a bioinformatics analysis to determine if the absence of helix α4 is a conserved property of this subfamily. The structure prediction program PsiPred (*43*) was used to evaluate the secondary structure of 84 receiver domains with cognate HWE-kinase partners. Of the 84 proteins considered, 71 were predicted to lack helix α4 (Figure 3-5A). A subset of 13 HWE-associated receiver domains are predicted to have helix α4 but then appear to lack the last helix α5 (Figure 3-5B). This suggests that most of the HWE-associated receiver domains in which helix α4 is present, have compensatory perturbations that disrupt helix α5 of the 455 face. By contrast, all of the α-helices and β-strands of the consensus Rossman fold are accurately predicted by the PsiPred program for a control group of 8 receiver domains of known structure that are not associated with HWE-kinases (Figure 3-5C).

Sma0114 Binds Mg^{2+} and Ca^{2+}

We next wanted to see if divalent cations, which are required for the stabilization of the acyl-phosphate linkage (*4, 14*), are able to bind to Sma0114 in spite of the differences seen for the 455 face. Several groups have used Mg^{2+} to study the metal-bound states of receiver domains (*11, 13, 44*). In the case of Sma0114 we found that Mg^{2+} binds very weakly, with a K_d near 75 mM and saturation at 150 mM Mg^{2+}. The $MgCl_2$ concentration required to stabilize the fully metal bound state leads to a decrease in sensitivity for NMR experiments due to high solution ionic strength, especially for data

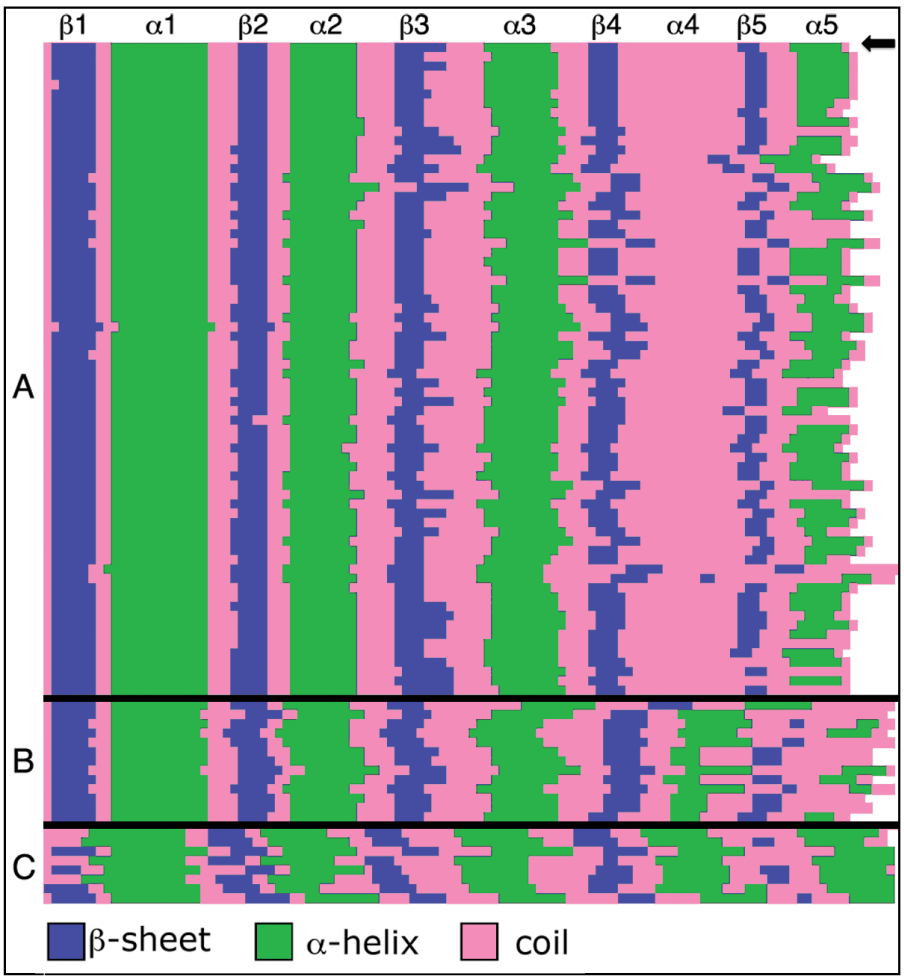

Figure 3-5 Secondary structure prediction of 84 HWE-associated receiver domains. Each row corresponds to one receiver domain. (**A**) Subset of 71 receiver domains with cognate HWE-kinases, all of which are predicted to be missing helix α4. Sma0114 is at the top of the figure labeled with an arrow. (**B**) Subset of 13 HWE-associated receiver domains predicted to have helix α4. Note that most of the members of this subset are predicted to be missing the alternate helix α5 of the 455 face. (**C**) Control group of 8 receiver domains of known α_5/β_5 structures used to verify the secondary structure prediction algorithm.

collection with a cryogenic probe. An additional problem is that the large Mg^{2+} concentration needed to saturate Sma0114 causes non-specific binding to clusters of acidic residues in the protein. Finally, Sma0114 is more susceptible to aggregation on the timescale of days to weeks needed for a detailed NMR characterization, in the presence of high (> 100 mM) concentrations of $MgCl_2$.

Due to the problems associated with Mg^{2+}, we carried out an NMR titration using $CaCl_2$ as an alternative metal. Superimposed ^{1}H-^{15}N spectra of apo and Ca^{2+}-bound Sma0114 are shown in Figure 3-6A. In contrast to Mg^{2+} which has a K_d of 75 mM (Figure 3-6B), Sma0114 binds Ca^{2+} with a ~1 mM K_d, and saturation is achieved at 1.5 mM concentrations of $CaCl_2$ (Figure 3-6C). Residues that show the greatest chemical shift changes upon addition of Ca^{2+} are indicated with boxes in Figure 3-6A, and for the most part cluster to the active site.

Figure 3-7 shows the chemical shift perturbations observed in ^{1}H-^{15}N HSQC spectra of Sma0114 due to addition of $CaCl_2$ (Figure 3-7A) and $MgCl_2$ (Figure 3-7B). The unusually large concentration of $MgCl_2$ necessary to achieve saturation induces non-specific binding to acidic Asp and Glu residues as illustrated by the large number of the chemical shift perturbations and line broadening effects observed (Figure 3-7B). By contrast, with $CaCl_2$ perturbations are mostly restricted to the active site of the enzyme (Figure 3-7A). Some residues become broadened rather than experiencing chemical shift changes upon addition of $CaCl_2$ or $MgCl_2$, and these are indicated with arrows in Figure 3-7. They include Glu17, which is one of the ligands for Ca^{2+}, and Lys105, which is part of the active site. Although most studies of metal-bound receiver domains have used Mg^{2+}, Ca^{2+} was used for structural studies of Spo0A and PhoP (*45, 46*).

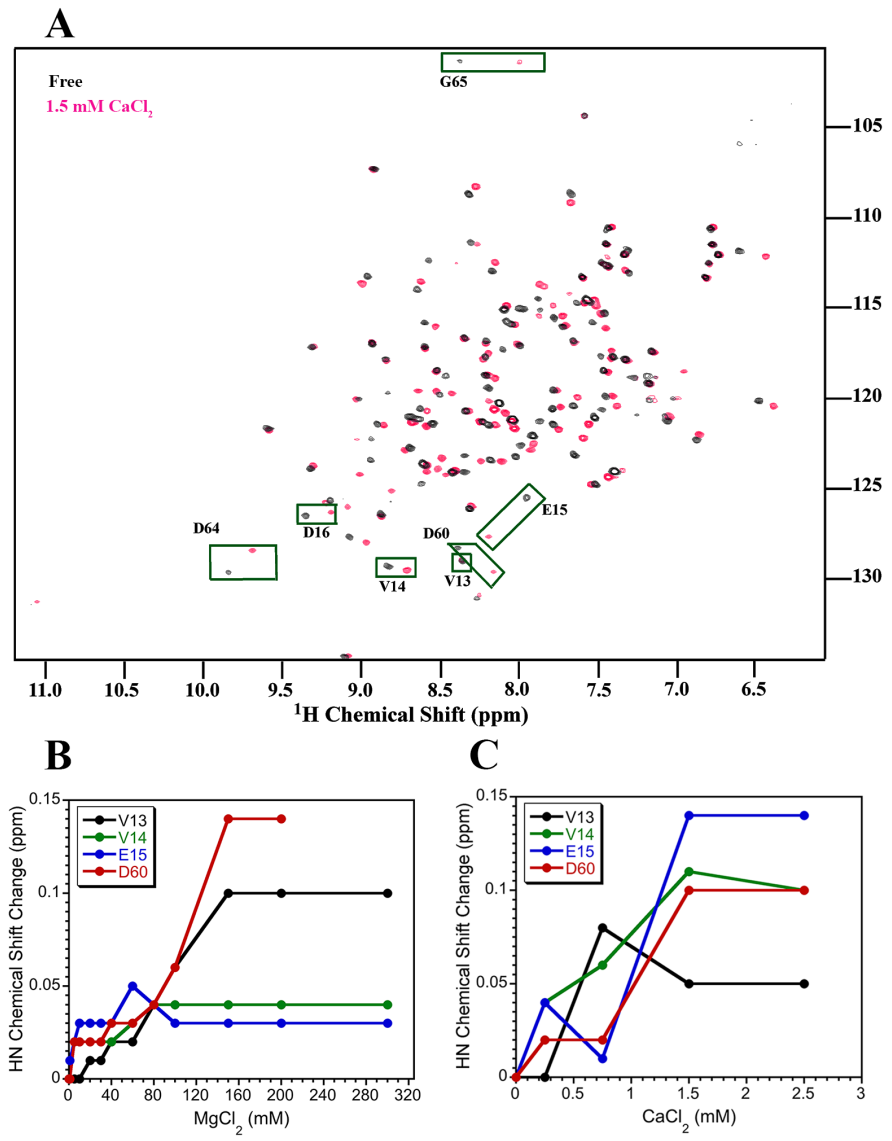

Figure 3-6 Metal-binding of Sma0114 followed by NMR spectroscopy. (**A**) Superposition of ^{1}H-^{15}N-HSQC spectra for the apo (black) and Ca^{2+}-bound forms of Sma0114 (pink). Active site residues that show the largest chemical shift perturbations and are indicated with green boxes. Titration curves for residues as a function of increasing $MgCl_2$ (**B**) or $CaCl_2$ (**C**) concentrations. Please note the different scaling of the x-axis in (**B**) and (**C**). Errors in chemical shift differences estimated from the digital resolution of the spectra were 0.02 ppm for (**B**) and (**C**).

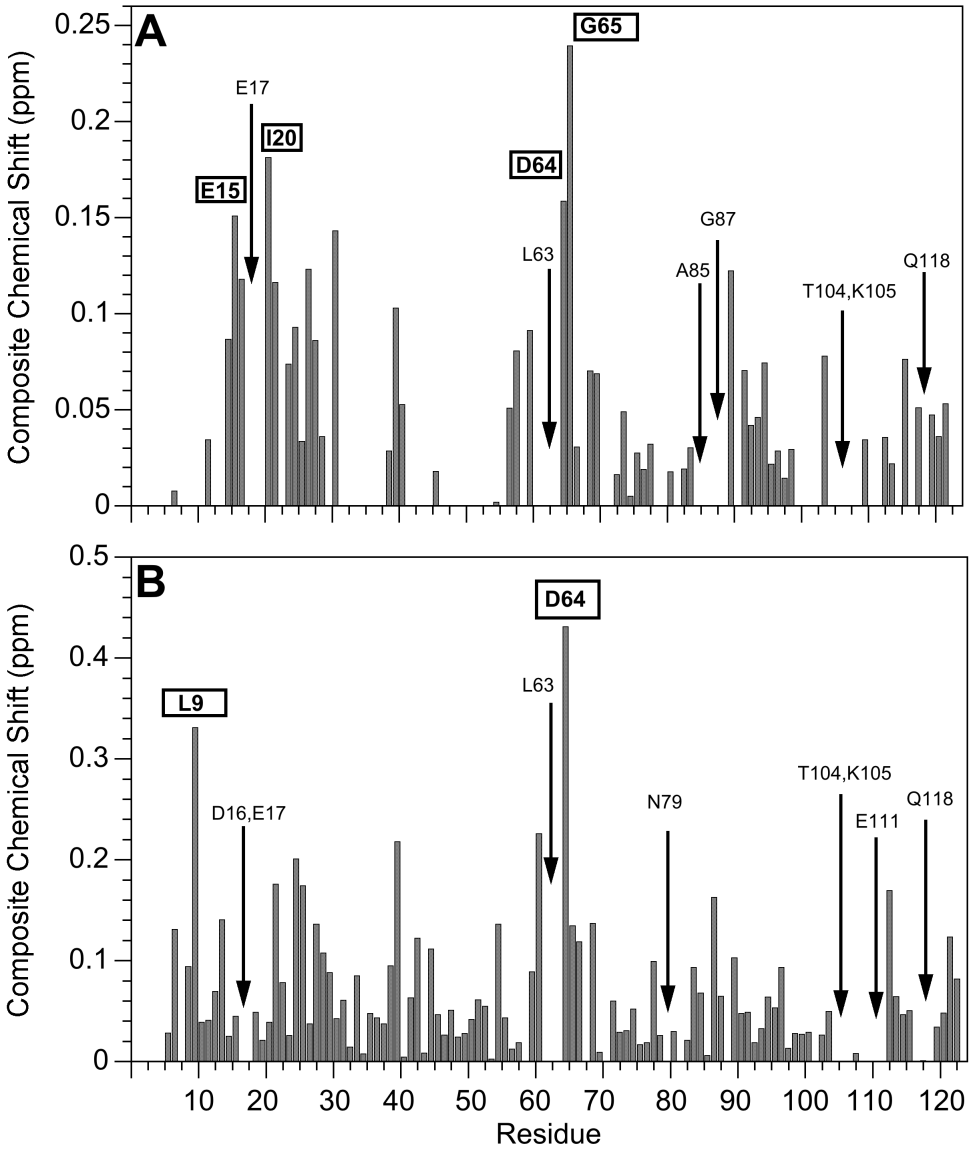

Figure 3-7 Perturbations of Sma0114 ^{1}H-^{15}N resonances due to metal-binding. Bar graphs show composite HN and N chemical shift changes ($\Delta HN + 0.1(\Delta N)$) that occur in the presence of 1.5 mM $CaCl_2$ (**A**) or 150 mM $MgCl_2$ (**B**). Residues in boxes indicate amino acids that experience large chemical shift perturbations upon metal binding. Arrows represent residues that broaden beyond detection in the presence of the metals. Errors in the composite chemical shift changes estimated from the digital resolution of the spectra were 0.03 ppm.

Discussion

Several groups have characterized the inactive and active conformations of receiver domains (*11, 39, 44*). Understanding the structural features of these enzymes has provided a wealth of insight into their functions. The Sma0114 receiver domain, which may be prototypical of receiver domains associated with HWE-kinases, has a conserved active site but shows substantial differences in the 455 face, which is used to propagate signal transduction intracellularly.

The structural and dynamic differences in the 455 face of Sma0114 suggest that this enzyme has an altered recognition interface for binding to downstream effectors. The PFxFATGY motif that is specific to receiver domains associated with HWE-kinases, forms strand β4 in the Sma0114 structure (Pro81-Thr86) and the rigid part of the following loop (Gly87 and Tyr88). The last residue in the PFxFATGY motif demarcates the start of the Tyr88-Leu93 segment, which replaces helix α4 in Sma0114, and shows low S^2 order parameters consistent with a dynamically disordered region. The only other receiver domain for which S^2 order parameters have been described is the inactive form of Spo0F (*19*). In the case of Spo0F, it was shown that except for the chain termini, the protein is rigid on the ps-ns timescale. In typical receiver domains, helix α4 is a key element of the 455 face that undergoes concerted structural rearrangements following phosphorylation to form a new binding surface for downstream effector proteins (*11, 40*). Not only is helix α4 missing in Sma0114 but secondary structure prediction analysis suggests that the absence of helix α4, or more rarely helix α5, is a conserved feature of HWE-associated receiver domains. A possible role for flexibility in the segment

corresponding to helix α4 is to allow binding of downstream effectors through an induced fit mechanism, thereby controlling the specificity of molecular association.

Residues Pro81 through Gly87 of the PFxFATGY motif, which form strand β4 and the beginning of the following loop, while rigid on the ps-ns timescale have raised $R2_{ex}$ contributions consistent with dynamics on the μs-ms timescale. The exchange contributions to R2 relaxation for these sites, suggest that they experience the dynamic conformational equilibrium between the inactive and active substates of the enzyme that is a hallmark of receiver domains (*4, 39*). The PFxFATGY sequence houses the conserved Thr86 of the Y-T coupling mechanism. The partner aromatic residue for Y-T coupling in Sma0114 is replaced by Leu103. In a previous study of the receiver domain CheY, it was shown that the β4/α4 loop and the N-terminal portion of helix α4 constitute the binding site for downstream effectors (*47*). Because part of the PFxFATGY motif in Sma0114 occupies the same position in the structure as the β4/α4 loop region in CheY, it is likely that Sma0114 has a similarly reduced binding site for downstream effectors. We initially thought that the PFxFATGY sequence motif in Sma0114, may thus substitute or circumvent the Y-T coupling mechanism typically used in receiver domains. NMR analysis of the activated enzyme was necessary to explore this hypothesis in Chapter 4.

In spite of the differences in the 455 face, Sma0114 retains the ability to bind divalent metal ions needed for activation. The residues in Sma0114 that experience the largest chemical shift perturbations with increasing metal concentrations are consistent with the binding site predicted based on sequence homology to other receiver domains (Figure 3-6).

Reprinted from Sheftic et al. (2013) (*48*)
I performed all experiments

Chapter Four. Characterization of Activated Sma0114

Abstract

Sma0114 is the first structurally-characterized representative from the family of receiver domains that are substrates for HWE-kinases. Presented here is the NMR structure of Sma0114 bound by Ca^{2+} and BeF_3^-, a phosphate analog that stabilizes the activated state. Differences between the NMR structures of the inactive and activated states occur in helix α1, the active site loop that connects strand β3 and helix α3, and in the segment from strand β5 to helix α5 of the 455 face. In Sma0114 the structural changes accompanying activation result in a more negatively charged surface for the 455 face. Coupling between the 455 face and active site phosphorylation is usually mediated through Y-T coupling. NMR analysis indicates that Sma0114 lacks Y-T coupling, and that communication between the active site and the 455 face is achieved through a conserved lysine residue that stabilizes the acyl phosphate in receiver domains. ^{15}N-NMR relaxation experiments were used to investigate the backbone dynamics of the Sma0114 the binary Sma0114•Ca^{2+} complex, and the ternary Sma0114•Ca^{2+}•BeF_3^- complex to complement the analysis of the apo protein. The loss of entropy due to ligand binding at the active site is compensated by increased flexibility in the 455 face. The dynamic character of the 455 face in Sma0114, which results in part from the replacement of helix α4 by a flexible loop, may facilitate induced-fit recognition of target molecules.

Introduction

In order to gain insight into the changes in Sma0114 induced by activation we solved the NMR structure of the enzyme in the activated state. Due to the fact that the lifetimes of acyl phosphates (~ 30 min) are too short to allow detailed NMR studies, we used the phosphate mimetic beryllium trifluoride (BeF_3^-) to stabilize the activated state, as has been done for other structural studies of receiver domains (*49, 50*). The structures of inactive and activated Sma0114 were analyzed for key differences that could communicate the phosphorylation state of the receiver domain from the active site to the 455 face of the protein. We compared the conformational changes accompanying activation of Sma0114 to three homologous single-domain receiver domains, to see what features are conserved. Complementary ^{15}N relaxation experiments were used to investigate how the dynamics of Sma0114 are affected by binding Ca^{2+}, and by activation with Ca^{2+} and BeF_3^-.

Results

NMR Structure of Activated Sma0114

Backbone (C^{α}, N, C') traces for the 26 lowest energy NMR structures of the Sma0114•Ca^{2+}•BeF_3^- ternary complex are shown in Figure 4-1A. Statistics pertaining to the quality of the NMR structures are given in Table 4-1. The precision of the activated NMR structure over elements of regular secondary structure is 0.71 Å, slightly better than the 0.86 Å precision of the inactive state (*31*). Other measures of the quality of the structure including the Procheck Z-score, RMS deviations from ideal geometry, and average residual NOE violations are also slightly better for the activated than inactive structure (*31*).

A cartoon showing the α_4/β_5 Rossmann fold topology of the activated structure is shown in Figure 4-1B. As previously described for the structure of the inactive state in Chapter 3, Sma0114 differs from typical receiver domains in having the fourth α-helix of the 455 face replaced by a flexible loop. Interestingly, the fourth helix is also absent in the structure of the activated state (Figure 4-1B). Thus binding of Ca^{2+} and BeF_3^- does not induce folding of the fourth α-helix.

Figure 4-1C shows the active site, and key residues that interact with the Ca^{2+} and Be_3F^- ions. The three residues Glu15, Asp16, and Glu17 in the loop between strand β1 and helix α1 ($L_{\beta 1\alpha 1}$) each have one oxygen within bonding distance of the Ca^{2+} ion (2.1, 2.1, and 2.4 Å). The Ca^{2+} ion is within bonding distance of two fluorine atoms from the BeF_3^- molecule (1.8 and 2.1 Å).

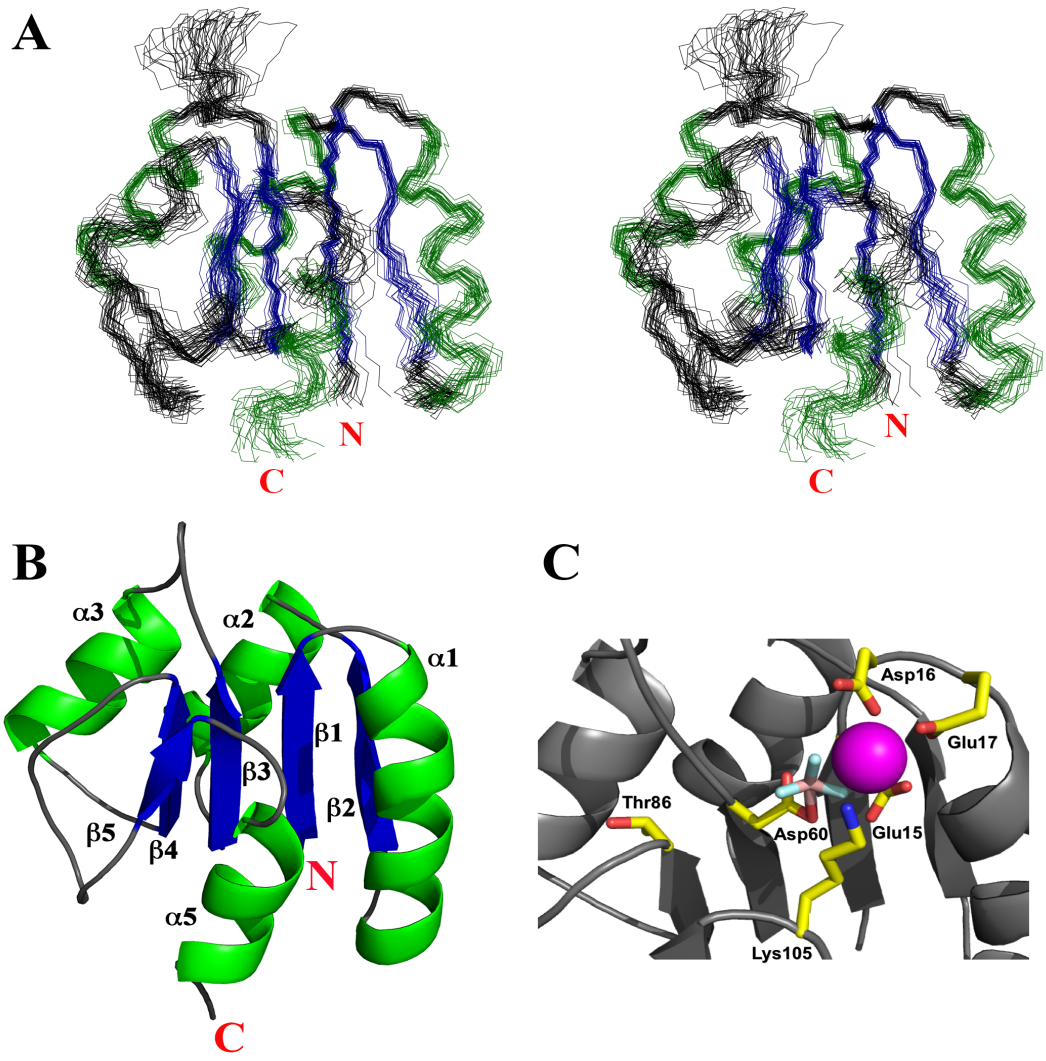

Figure 4-1 NMR Structure of the ternary Sma0114•Ca^{2+}•BeF_3^- complex. (**A**) Stereo-diagram of the 26 lowest energy NMR structures superimposed over regular main-chain secondary structure elements: β1 (10-15), α1 (18-31), β2 (34-39), α2 (42-51), β3 (55-60), α3 (70-78), β4 (81-86), β5 (102-104), α5 (110-118). (**B**) Backbone cartoon of the NMR structure closest to the ensemble mean. (**C**) Sidechains that participate in ligand binding at the active site of Sma0114 and Thr86. The Ca^{2+} ion is shown in purple and the BeF_3^- molecule in pink and cyan.

Table 4-1: Statistics for the 26 Lowest Energy NMR Structures of

Sma0114•Ca^{2+}•BeF_3^-

NMR restraints (total)	1791		
Distance (total)	1644		
Intraresidue NOEs	472		
Sequential NOEs	479		
Short range NOEs (1< |i-j| < 5)	136		
Long range NOEs (5 ≤ |i-j|)	495		
Hydrogen bonds (31x2)	62		
Dihedral (ϕ 70, ψ 61, χ_1 16)	147		
Residual restraint violations			
NOE (Å)[a]	0.033 ± 0.0004		
Dihedral (°)[b]	0.587 ± 0.075		
RMS deviations from ideal geometry			
Bonds (Å)	0.0037 ± 0.0002		
Angles (°)	0.724 ± 0.004		
Improper torsions (°)	0.560 ± 0.007		
van der Waals energy (kcal/mol)[c]	87.85 ± 1.75		
Lennard-Jones energy (kcal/mol)[d]	-85.42 ± 6.33		
ProcheckNMR Z-score[e]	-2.11		
Ramachandran statistics for all residues[e] (and excluding loops[f])			
most favored (%)	58.8	(58.2)	
allowed (%)	29.6	(36.8)	
generously allowed (%)	7.5	(4.9)	
disallowed (%)	4.1	(0.0)	
Coordinate RMS deviations (Å)			
NMR ensemble to average	C_α, C, N		All Heavy
Entire domain (110 residues)[g]	0.88		1.34
Excluding loops (70 residues)[f]	0.71		1.19
NMR model to inactive model			

Entire domain (110 residues)[g]	1.91	2.70
Excluding loops (70 residues)[f]	1.66	2.32

---

[a] Structures have no NOE violations greater than 0.3 Å
[b] Structures have no dihedral violations greater than 5°
[c] E_{vdw} was calculated using the X-PLOR *F*repel function (*34*) with van der Waals interactions and atomic radii set to 0.8 times their CHARMM (*35*) values.
[d] *E*L-J was calculated using the CHARMM empirical energy function (*35*).
[e] ProcheckNMR used via Protein Structure Validation Suite (PSVS) (*36*).
[f] Only residues in regular secondary structure (10-15, 18-31, 34-39, 42-51, 55-60, 70-78, 81-86, 102-104, 110-118).
[g] Excluding N-terminus (residues 1-9) and C-terminus (residues 119-123).

The third fluorine atom does not appear to form any bonding interactions. The fluorine atom closest to the Ca^{2+} is 2.6 Å away from the N_z atom of Lys105. The positively charged N_z atom of Lys105 may also be involved in an electrostatic interaction with the OE2 atom of Glu15, with the two separated by 3.9 Å. The Be atom of BeF_3^- is bonded to the OD2 atom of Asp60 (2.0 Å).

Conformational Changes Induced by the Activation of Sma0114

The structures of Sma0114 in the inactive and activated states, modeled by the Sma0114•Ca^{2+}•BeF_3^- ternary complex, are shown superimposed in Figure 4-2A. A plot of the Cα RMSD after superimposition of the two structures is shown in Figure 4-2B. The mean RMSD over all residues is 1.91Å (Table 4-1). Regions of the protein that show larger differences than the mean RMSD are colored in pink and labeled (Figure 4-2A).

The largest difference is seen for the loop between strand β3 and helix α3 ($L_{\beta3\alpha3}$) that partially covers the active site in the inactive structure but is in an open conformation in the activated state. As expected, differences are seen in the 455 face. The C-terminus of strand β5 becomes shorter by one residue, Lys105, which becomes part of $L_{\beta5\alpha5}$ in the activated state. The C-terminus of strand β5 moves away from the core of the structure by about 1.0-1.5 Å in the activated state, and undergoes a twist of about 30 degrees to align better with strand β4. In the activated state, helix α5 undergoes a tilt of about 10-30 degrees to make it less parallel with the β-strands compared to the inactive structure.

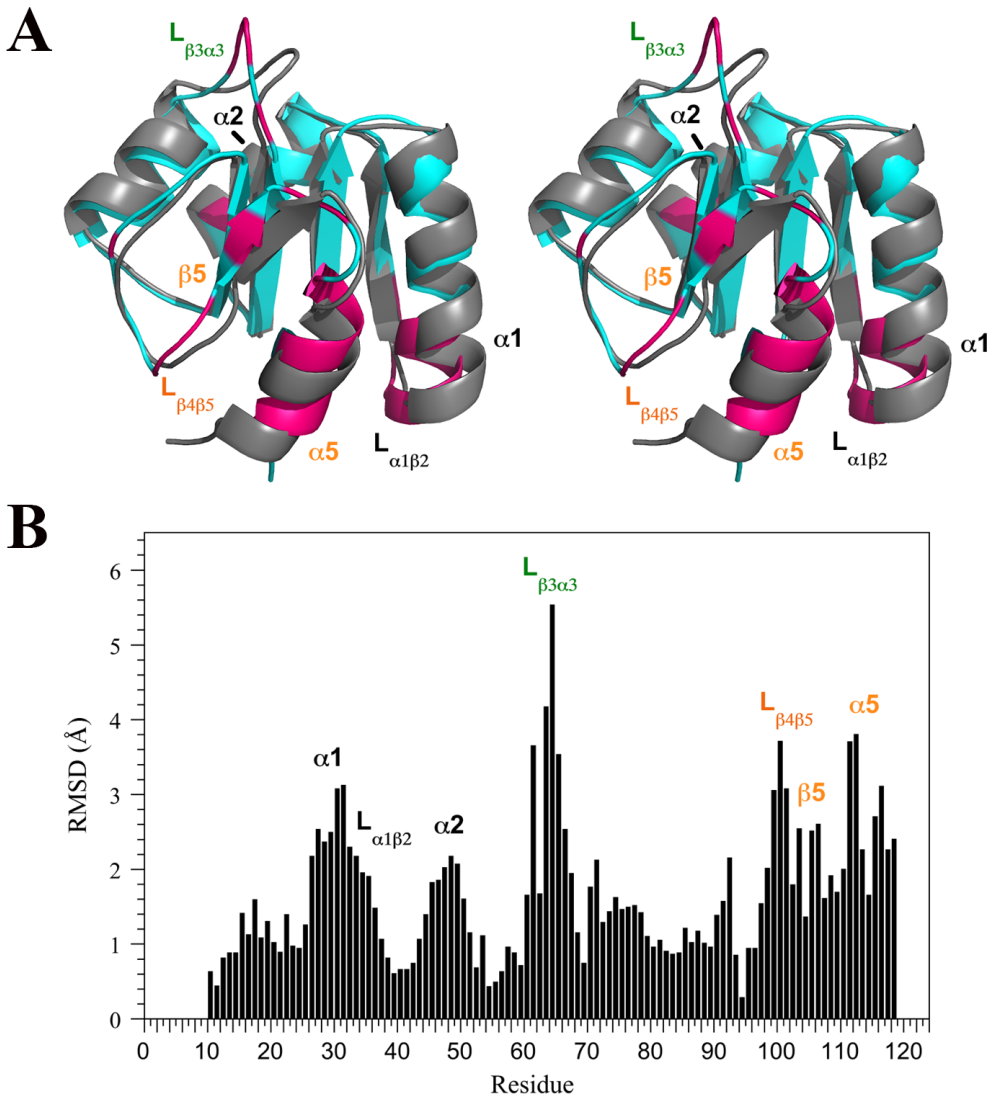

Figure 4-2 Comparison of the inactive and activated NMR structures of Sma0114. (**A**) Stereo-diagram showing the least-squares superimposition of the inactive (grey) and activated (cyan) structures. Residues with the largest RMS differences between the two structures are indicated in pink on the activated structure. (**B**) Bar plot of Cα RMSD values between the inactive and activated structures. Labels indicate the largest differences between the two structures, with green and orange labels indicating segments from the active site and 455 face, respectively.

The helix also undergoes translational displacements in the activated state, including a shift of 1.0 Å towards the top of the molecule (defined by the arrows indicating the C-termini of the β-strands in Figure 4-2A) and a move of about 1.5 Å closer to the core of structure. Outside of the 455 face, there are changes in α-helices 1 and 2 (Figure 4-2). Helix α1 undergoes a bend in the activated structure which brings the last two turns about 1-2 Å closer to helix α5 and to the core of the structure. Similarly, helix α2 experiences a 10-20 degree axis tilt in the activated state, that brings the C-terminus of the α-helix closer to the β-sheet core of the protein. Overall, the movement of α-helices 1 and 5 towards each other and the protein core, make the activated state structure slightly more compact than the inactivated structure. The radii of gyration calculated from the structures are 14.2 Å for the inactive and 13.8 Å for the activated states.

Comparison to Other Receiver Domains

To investigate the extent to which the changes associated with activation of Sma0114 are conserved in other receiver domains, we looked at three single-domain response regulators NtrC, CheY and Spo0F, shown in Figure 4-3. As in Figure 4-2, the largest changes between the inactive and activated states are highlighted in pink on the activated structures (Figure 4-3A) and bar graphs are used to summarize Cα RMSDs (Figure 4-3B). The analysis shows that the conformational changes induced by the activation of receiver domains can be quite variable.

In NtrC and CheY, conformational changes occur primarily along the 455 face of the activated enzymes, particularly for helices α4 and α5.

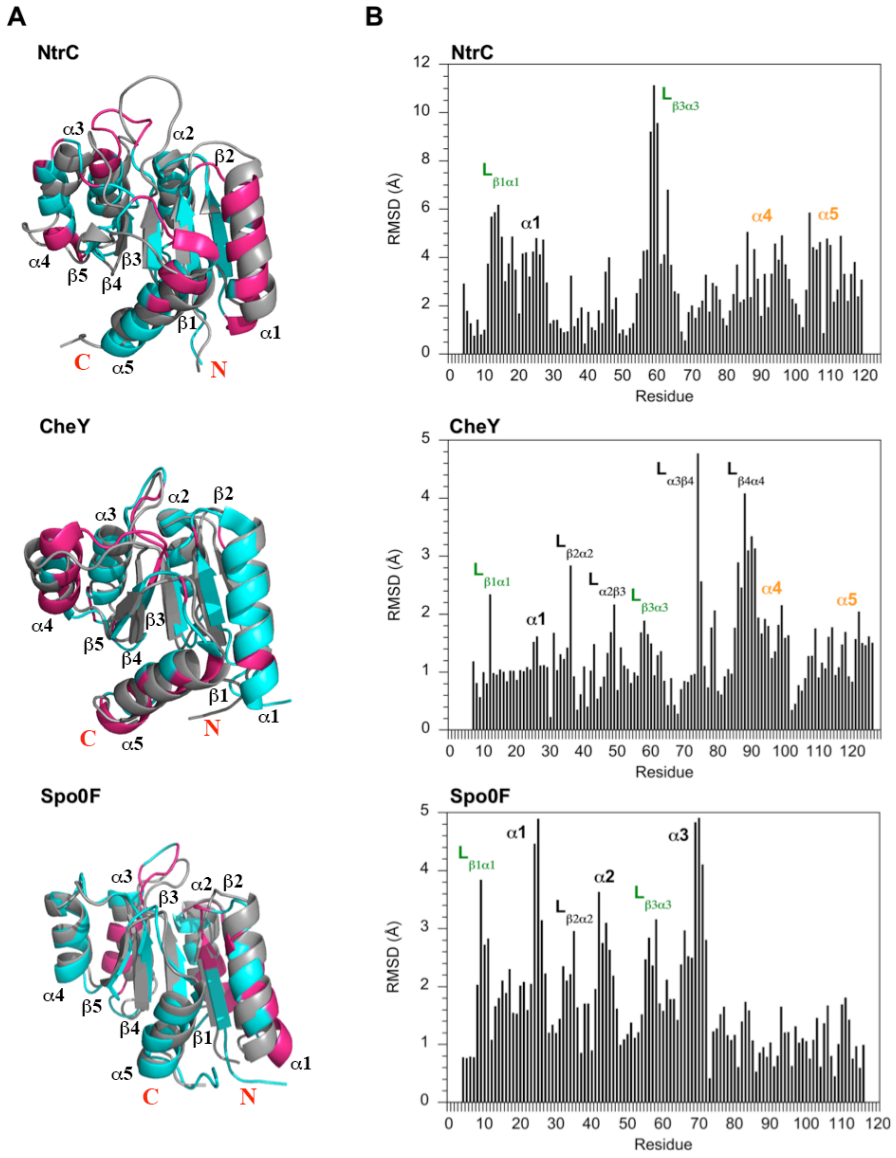

Figure 4-3 Structurally homologous single domain response regulators. (**A**) Superimposition of inactive (grey) and activated (cyan) structures of NtrC (PDB codes 1NTR and 1KRX), CheY (PDB codes 5CHT and 1DJM) and Spo0F (PDB codes 1NAT and 1PUX). Regions that show the largest differences upon activation are colored in pink on the activated structures. (**B**) Corresponding Cα RMSD plots showing the largest differences between the inactive and activated structures.

Despite the absence of helix α4 in Sma0114 we do observe changes in the structure of the β5 - α5 segment (Figure 4-2). The Spo0F receiver domain shows only small differences for the 455 face, with the majority of structural changes occurring for the first three α-helices outside of the 455 face (Figure 4-3) (*38, 39*). Like other single domain response regulators shown in Figure 4-3, Sma0114 (Figure 4-2) shows structural changes for helix α1 between the inactive and active states. This suggests that helix α1 could participate together with the 455 face in binding downstream effectors, or alternatively that it may be important in the transition to an activated state. Indeed, mutagenesis experiments indicate that helix α1 of receiver domains plays a role in determining the kinase specificity of receiver domains (*51*). The types of structural changes that occur upon activation appear to follow similar trends. As in Sma0114, activation of the NtrC, CheY, and Spo0F single domain receiver domains brings the α-helices closer to the β-sheet core of the structures (Figure 4-3A). Another conserved theme is that helices α1 and α5 move closer together in the activated states (Figure 4-3A).

The largest changes observed between the inactive and activated states are in the active site loops, particularly $L_{\beta3\alpha3}$ following the phosphorylation site (Figure 4-3B). In all three structures, except CheY, where the differences are relatively small, $L_{\beta3\alpha3}$ moves away from the active site in the activated state (Figure 4-3A). The open conformation of $L_{\beta3\alpha3}$ was seen for NtrC (*11*) but the authors noted that the precision of that region in the inactive structure precluded additional analysis. In Sma0114 the average RMSD for residues comprising $L_{\beta3\alpha3}$ (Val61-Glu66) between the inactive and activated states is 3.2 Å. These RMSD values are much larger than the internal precisions of the structures

(RMSDs of 0.94 for the inactive (Table 3-1) and 0.88 for the activated structure (Table 4-1)). That the displacement of $L_{\beta 3 \alpha 3}$ appears to be a conserved theme for the activated states of receiver domains suggests that its open conformation plays an important role in activation.

Sidechain Changes Accompanying Activation

An interesting question is how the phosphorylation state of the activated enzyme is communicated to the 455 face in order to achieve the conformational transition that prepares the receiver domain to bind its cognate effector proteins. Moreover, since binding to partner molecules is likely to involve surface interactions, it is important to characterize how activation affects the surface of the receiver domain.

In typical receiver domains, active site phosphorylation propagates structural changes to the 455 face through Y-T coupling (*1*). Upon phosphorylation of a conserved aspartate, a hydrogen bond forms via a Thr/Ser sidechain hydroxyl with the phosphate. The displacement of a Thr/Ser leaves an internal cavity that is filled by the reorientation of an Tyr/Phe residue from the surface to the interior of the structure (*1*). Thus, the coupled rearrangement is the conformational change linking the active site and output 455 face. In Sma0114 the conserved threonine corresponds to Thr86. As shown in Figure 4-4A, Thr86 in strand β4 is in a nearly identical position before and after activation. The hydroxyl oxygen of Thr86 is 7.0 Å away from the closest fluorine atom in BeF_3^-, and is therefore too far to form a bond. Leu103 in strand β5 replaces the partner aromatic residue for the Y-T coupling mechanism commonly seen in receiver domains.

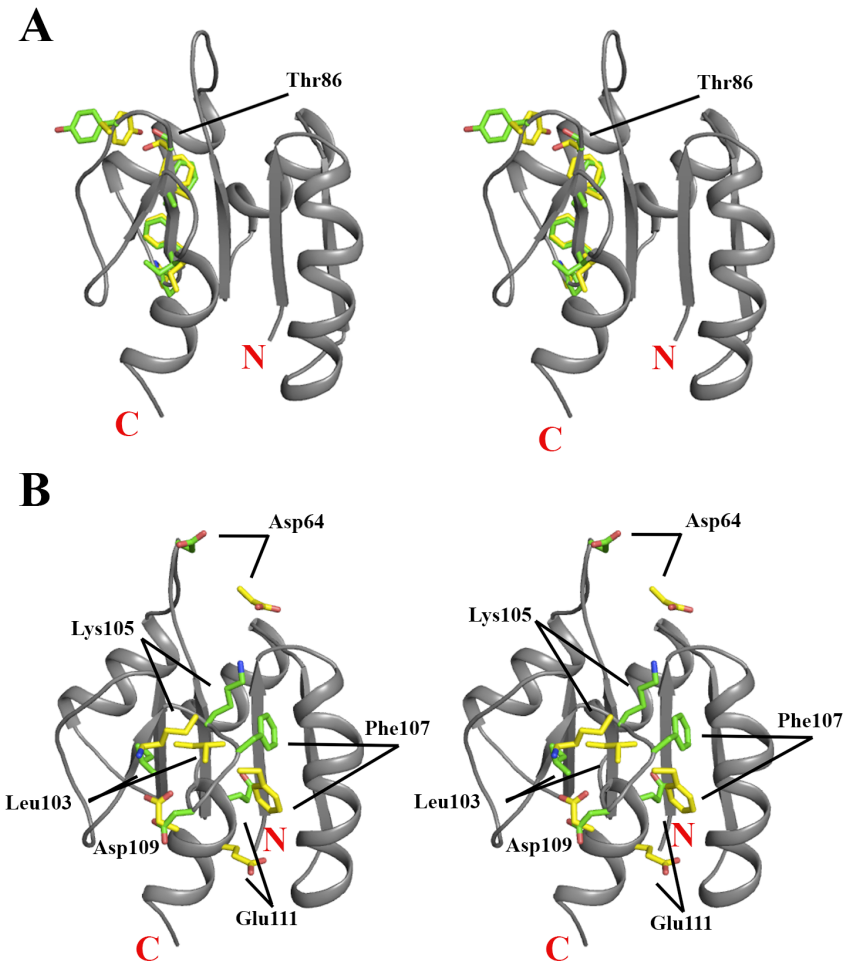

Figure 4- 4 Stereo-diagrams comparing selected sidechains between the inactive and activated states of Sma0114. (**A**) The sidechain conformations of PFxFATGY motif residues that show little change. (**B**) Sidechains that undergo large structural changes between the inactive and activated conformations include Asp64 at the active site, and Leu103, Lys105, Phe107 and Glu111 in the 455 face of the domain. The backbone trace structure is shown for the activated state (gray). Sidechains from the inactive structure are shown with yellow carbons while those from the activated state are shown with green carbons.

The sidechain of Leu103 does undergo a reorientation upon activation but from the interior of the protein to the surface, the opposite of what is seen for the corresponding aromatic residue in typical receiver domains. These observations suggest that like the CheY2 receiver domain (*37*), Sma0114 lacks the Y-T coupling mechanism.

Based on the structure of the inactive protein, we proposed that the Y-T coupling mechanism in Sma0114 may be substituted by the PFxFATGY sequence motif. This sequence motif is highly conserved in the subfamily of receiver domains that are substrates for HWE-kinases. As shown in Figure 4-4A, the PFxFATGY sequence motif, which includes Thr86 and forms strand β4 of the Sma0114 structure, shows only minor structural differences between the inactive and activated protein (also see Figure 4-2B). Of the residues in the PFxFATGY motif, Thr86 comes closest to the BeF_3^- molecule, through its Hα proton, which is 4.1 Å away. Given that there are no major structural changes in the PFxFATGY motif between the inactive and activated structure it seems unlikely that this motif plays a role in activation. Due to it's conservation in the HWE-kinase associated subfamily it is possible that the PFxFATGY motif has a role in kinase binding or is part of the modified surface of Sma0114 used to bind effector molecules.

Figure 4-4B shows the sidechains that experience the largest changes between the inactive and activated structures. Lys105 is one of the residues with the greatest perturbation. In the inactive state Lys105 forms a salt-bridge with Asp109 on the 455 surface of the domain. This salt bridge is broken in the activated structure, where the Lys105 sidechain enters the interior of the protein to bind BeF_3^-. The repositioning of Lys105 to the interior of the domain is accompanied by the displacement of Leu103 in strand β5 from the interior towards the surface of the structure. The cavity left by the

displacement of Leu103 is filled by Lys105 and by the aromatic ring of Phe107 from loop $L_{\beta5\alpha5}$. The movement of Phe107 leaves room for Glu111 to switch from a position on the surface exposed side of helix $\alpha5$ to one that bridges the gap between helices $\alpha5$ and $\alpha1$. Part of what may contribute to bringing helix $\alpha1$ closer to the β-sheet core of the protein is an electrostatic interaction between His33 NE2 and Asp55 OD1, which are 4.3 Å apart in the activated state versus 8.2 Å apart in the inactive structure.

Outside of the 455 face, large changes in sidechain positions are found at the active site. These include Glu17, which moves from a surface position towards the interior of the structure to act as a ligand for the Ca^{2+} ion. Large changes are also seen for the entire loop $L_{\beta3\alpha3}$, spanning residues Val61-Glu66, which moves away from the active site to an 'open' conformation in the activated structure (Figure 4-4B). The displacement of $L_{\beta3\alpha3}$ in the activated structure may be due to steric occlusion from the BeF_3^- molecule, Ca^{2+} ion, and the ligating residues from the protein that move into the active site. Electrostatic repulsion between Asp64 at the center of this loop and the high density of negatively charged groups in the active site (Glu15, Asp16, Glu17, Asp60) may also contribute to the displacement of the loop.

The sidechain changes accompanying activation of Sma0114 lead to changes in the surface properties of the 455 face, as shown in Figure 4-5. Sma0114 has a predominantly anionic character with a pI of 4.5, due to an excess of 21 acidic over 11 basic residues. Most of the surface of Sma0114 is negatively charged, except for a positively charged patch corresponding to the exposed sides of helices $\alpha2$ and $\alpha3$.

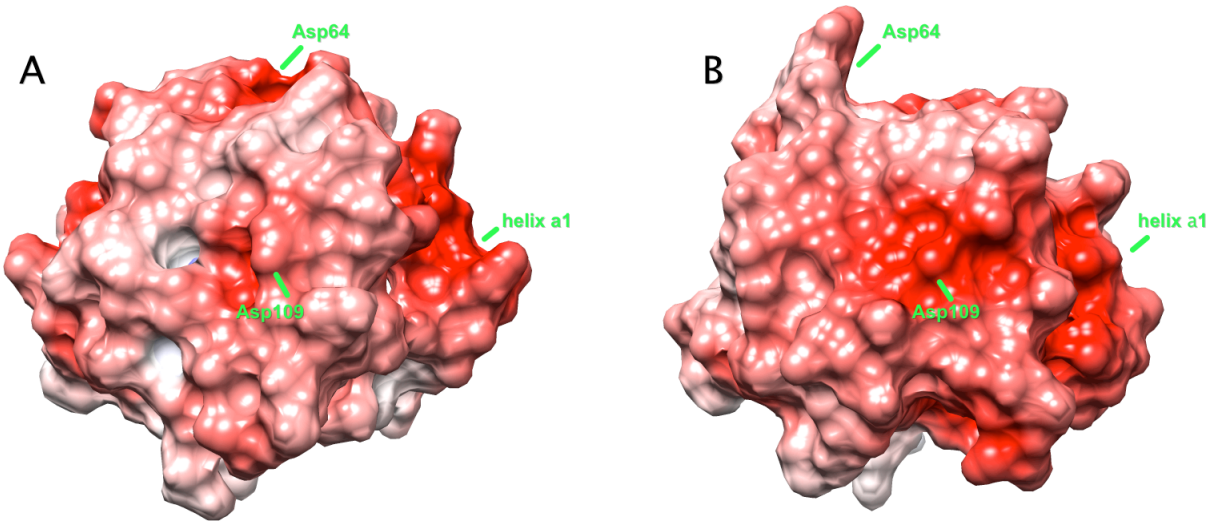

Figure 4-5 Electrostatic potential surface for the 455 face in (**A**) the inactive domain and (**B**) the activated domain. Red and blue correspond to negative and positive electrostatic potentials, respectively. The positions of Asp64, Asp109, and helix α1 are indicated on the surfaces. Electrostatic potentials were calculated with the APBS server (*52*).

This unusual charge distribution is unique to Sma0114, and is not a feature shared with other receiver domains. The views in Figure 4-5 show the 455 face in front. The ridge to the right below the 455 face, is due to helix $\alpha 1$. The interface between helices $\alpha 1$ and $\alpha 5$ that closes the Rossmann fold structure has a high density of negatively charged residues. Activation of Sma0114 appears to increase the negative character of the 455 surface. This is particularly evident near Asp109, which is labeled in Figure 4-5. In the inactive state (Figure 4-5A) the charge on Asp109 is complemented by a salt bridge with Lys105. Upon activation Lys105 switches to the interior of the protein to bind BeF_3^-, leaving the negative charge on Asp109 exposed (Figure 4-5B). Other more subtle changes such as partial burial of the basic residues Lys91 and Arg96 also contribute to the display of a more negatively charged 455 surface in the activated state (Figure 4-5B).

Effects of Ca^{2+} and BeF_3^- Binding on Backbone Dynamics

Figure 4-6 shows S^2 order parameters, that describe the amplitudes of backbone HN motions on the ps-ns timescale, for inactive Sma0114 (Figure 4-6A) , the Sma0114•Ca^{2+} binary complex (Figure 4-6B), and the activated Sma0114•Ca^{2+}•BeF_3^- ternary complex (Figure 4-6C). The active site loops, $L_{\beta 1\alpha 1}$ and $L_{\beta 3\alpha 3}$ that contribute ligands to Ca^{2+} and BeF_3^-, show average S^2 order parameters. The lack of ps-ns motions in these loops indicate a lack of dynamic flexibility for these regions in the inactive state (Figure 4-6A). As we only have data on backbone dynamics, this does not exclude increased sidechain flexibility for these residues. In the inactive protein (Figure 4-6A), low S^2 order parameters indicative of backbone flexibility occur at the chain termini and

in the portion of loop $L_{\beta4\beta5}$ that corresponds to helix α4. The chain termini, and residues (Lys91, Gly92, Leu93) in the loop retain low S^2 terms in the all three states. Thus the region corresponding to helix α4 remains unfolded in the metal-bound and activated states.

Backbone flexibility is increased at a number of sites in the Sma0114•Ca^{2+} binary complex (Figure 4-6B) compared to the inactive protein (Figure 4-6A). This includes residues Met19 and Leu24 following the metal binding site in helix α1. In the 455 face increased mobility is seen for residues Leu102 in strand β5 and Val115-Val117 at the end of helix α5. The enhanced backbone dynamics in regions of the protein remote from the binding site, suggest an entropy-compensated binding mechanism (*53-56*), where the ordering of the Ca^{2+}-binding site is offset by increased mobility in other parts of the molecule.

Data on the backbone dynamics for the activated state, stabilized by the binding of Ca^{2+} and BeF_3^- are shown in Figure 4-6C. The S^2 order parameters of the inactive and activated protein are mapped on their corresponding structures in Figure 7. Mobility in helix α1 is reduced in the activated state (Figure 4-6C) compared to the Ca^{2+}-bound state (Figure 4-6B); except for residues Ile20 and Met22 immediately following the Ca^{2+} binding loop, which show intermediate order parameters (Figure 4-7). Like the Ca^{2+}-bound state, the activated state shows flexibility for residues in the 455 face, although there are some subtle differences compared to the former. In the activated state residues Thr104-Phe107 in loop $L_{\beta5\alpha5}$ show raised mobility whereas in the Ca^{2+}-bound state residue Leu102 in strand β5 is mobile (Figure 4-6B, 4-6C).

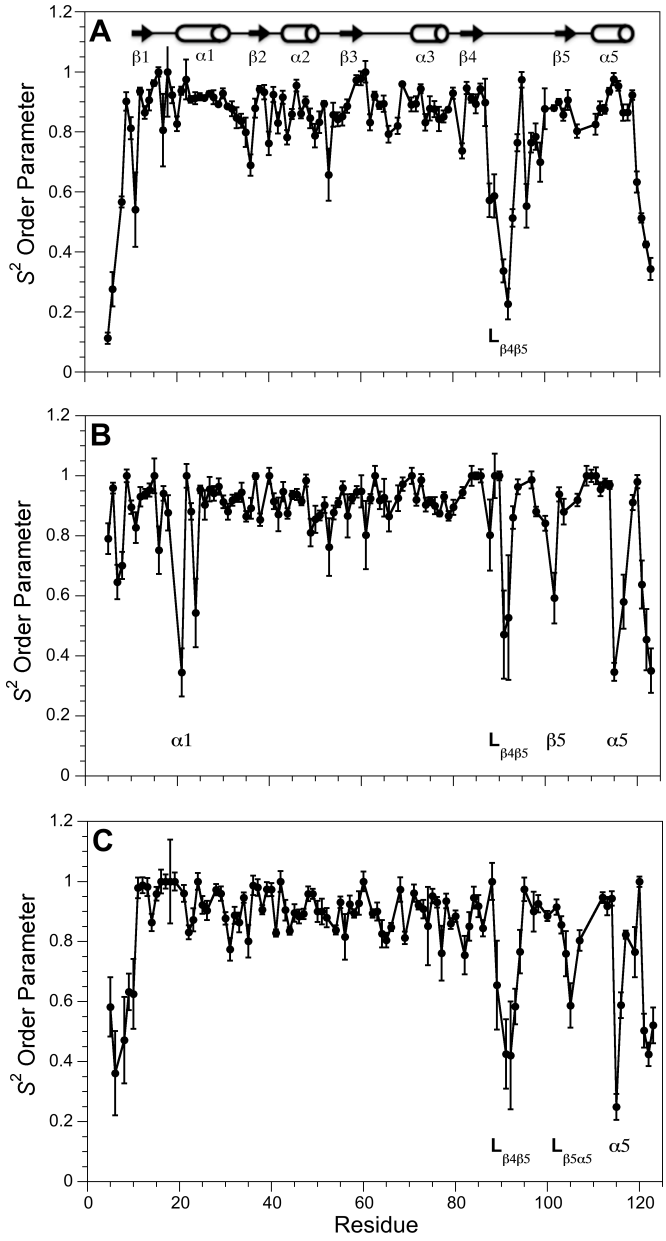

Figure 4-6 S^2 order parameters for Sma0114 in the apo state (**A**) the Ca^{2+} binary complex (**B**) and the Ca^{2+}•BeF_3^- ternary complex (**C**). The correlation times for each state were 5.6 ns for A, 5.9 ns for B, and 5.7 ns for C. A secondary structure diagram with α-helices represented by cylinders and β-sheets by arrows is shown in (A). Segments of the protein that show low S^2 order parameters are labeled in each panel.

Interestingly, the residue with the highest backbone mobility in loop $L_{\beta5\alpha5}$ of the activated state is Lys105, whose sidechain binds BeF_3^-. Like the Ca^{2+}-bound protein, the activated state exhibits low S^2 order parameters for residues Val115-Val117 at the C-terminal end of helix α5.

Figure 4-8 shows $R2_{ex}$ line-broadening contributions for the three states obtained from Model-Free calculations (*32, 42, 57*). In both the apo protein (Figure 4-8A) and in the Ca^{2+}-bound state (Figure 4-8B) there are four regions that have large $R2_{ex}$ terms indicative of conformational exchange on the μs-ms timescale. These are (I) the metal binding loop ($L_{\beta1\alpha1}$), (II) $L_{\beta3\alpha3}$ following the site of phosphorylation Asp60, (III) the C-terminal region of strand β4, and (IV) the β5-$L_{\beta5\alpha5}$ segment that holds Lys105, which stabilizes the acyl-phosphate moiety. $R2_{ex}$ contributions for these regions have been described in other receiver domains, and have been attributed to a dynamic equilibrium between the inactive and active conformations (*6, 20*). Consistent with previous work on other receiver domains (*6, 20*), the $R2_{ex}$ terms for regions I-IV in Sma0114 decrease as the equilibrium is shifted in favor of the activated state, by binding Ca^{2+} and BeF_3^- (Figure 4-8C). The large number of $R2_{ex}$ terms between 0 and 5 Hz for the activated state are an artifact of lower precision ^{15}N relaxation data. The relatively high salt concentration of 30 mM NaF, needed to generate BeF_3^- in the sample, reduces the sensitivity of the relaxation data.

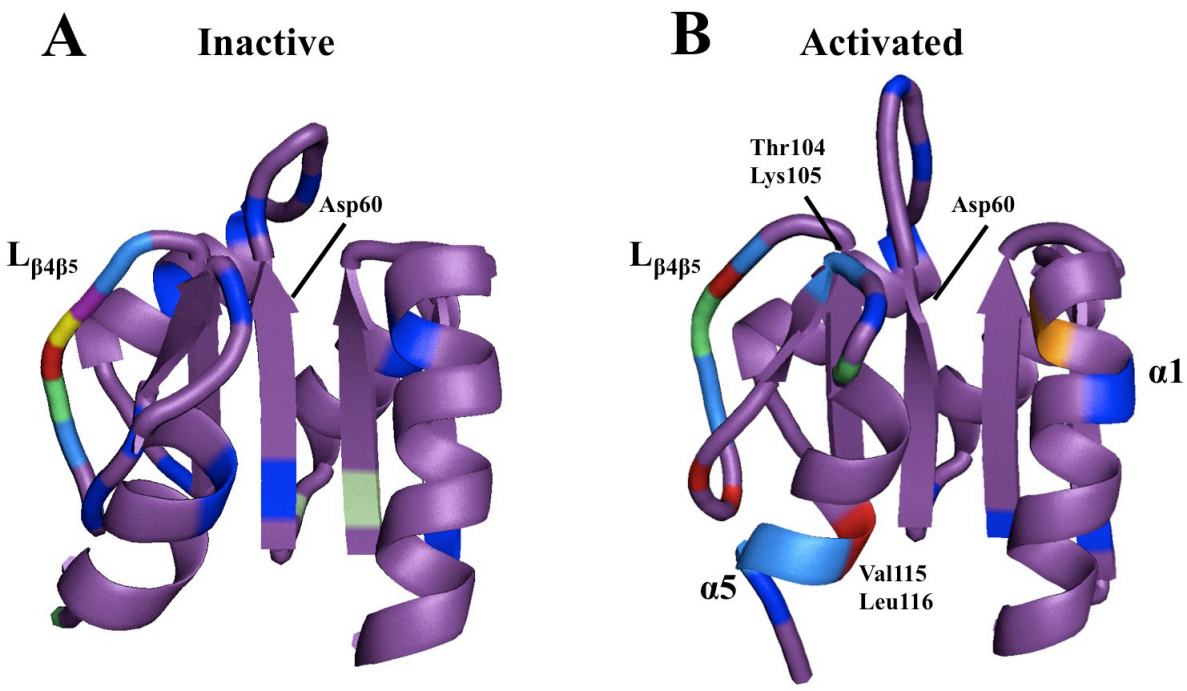

Figure 4-7 S^2 Order parameters mapped onto the NMR structures of (**A**) inactive and (**B**) activated Sma0114. The data are from Fig. 4-6A and 4-6C. The S^2 values are represented with a ROYGBIV seven-color gradient with red being the most flexible $0.00 \leq S^2 \leq 0.14$ and violet the least flexible $0.84 < S^2 \leq 1.00$. In the inactive structure, flexibility is predominantly associated with the loop $L_{\beta4\beta5}$ which replaces the canonical receiver domain fold helix α4 in Sma0114 (A). In the activated domain the loop $L_{\beta4\beta5}$ remains flexible and additional flexibility is seen for residues in helix α1, loop $L_{\beta5\alpha5}$, and helix α5, with the latter two belonging to the 455 face of the domain. Residues experiencing the greatest change in flexibility, Thr104, Lys105, Val115 and Leu116, are labeled on the activated structure. The phosphorylation site (Asp60) is labeled in (A) and (B).

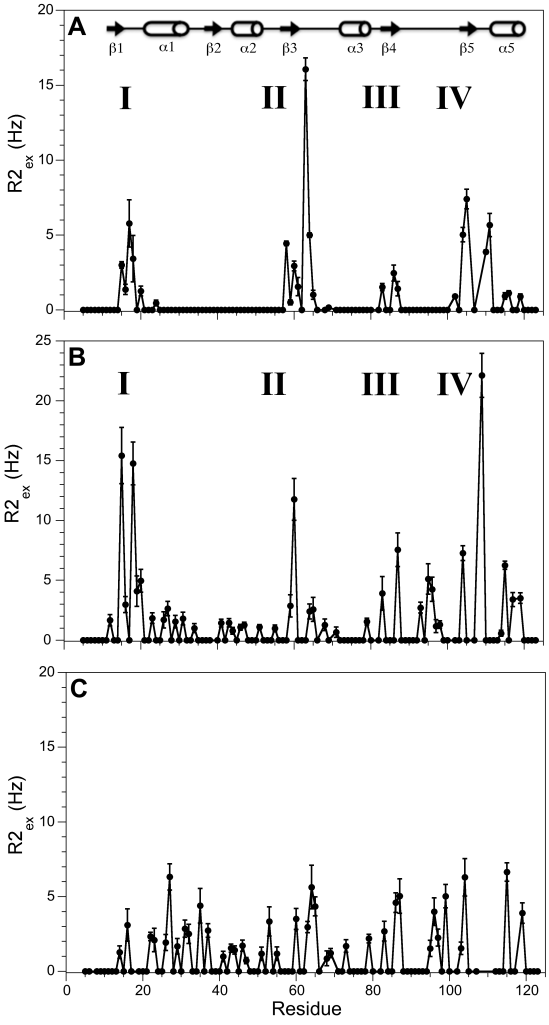

Figure 4-8 $R2_{ex}$ contributions for (**A**) apo-Sma0114 (**B**) the Sma0114•Ca^{2+} binary complex and (**C**) the Sma0114•Ca^{2+}•BeF_3^- ternary complex. The secondary structure is shown in (A). The apo domain (A) shows four regions with raised $R2_{ex}$ values: (I) the metal-binding site, (II) the phosphorylation site, (III) the C-terminal end of strand β4, and (IV) the junction between stand β5 and helix α5. The $R2_{ex}$ contributions persist when Ca^{2+} is bound (B) but are suppressed in the Sma0114•Ca^{2+}•BeF_3^- ternary complex.

Discussion

Phosphorylation causes a conformational switch that enables a receiver domain to bind downstream target molecules (*5*). The duration of the output response, which can range from seconds to hours, is regulated by the lifetime of the acyl phosphate linkage that induces the conformational transition to the activated state. In the current study we stabilized the activated state of Sma0114 by binding Ca^{2+} and the phosphate analog BeF_3^-, as has been done in other studies of receiver domains (*37, 50*) due to the limited lifetime of the acyl phosphate. In the present work we describe the NMR structure and dynamics of the activated form of Sma0114 to provide a more complete picture of the activation mechanism of this enzyme that is prototypical for the family of receiver domains activated by HWE-kinases.

A large difference between the structures of inactive and activated Sma0114 occurs in the loop $L_{\beta3\alpha3}$. In the inactive state this loop folds over the active site, partially occluding it (Figure 4-2A). Our active state structure shows that $L_{\beta3\alpha3}$ retains an open conformation in the activated state that leaves the active site more exposed compared to the inactive state (Figure 4-2A). One possible advantage of having $L_{\beta3\alpha3}$ displaced from the active site is that it could facilitate a water molecule entering the active site to perform hydrolysis of the phosphate group. The release of the phosphate-leaving group after the autophosphatase activity of Sma0114 would also be facilitated by the open $L_{\beta3\alpha3}$ that exposes the phosphorylation site Asp60. As such, loop $L_{\beta3\alpha3}$ could participate in determining the lifetime of the activated state before the enzyme switches back to its inactive state.

A key question in the structural biology of receiver domains is how phosphorylation of the conserved aspartate induces the conformational switch needed to bind downstream effector molecules. Typically the initial conformational switch occurs through Y-T coupling when a threonine at the C-terminal end of strand β4 repositions to hydrogen bond to the phosphate. The movement of the threonine allows the reorientation of an aromatic residue (typically a tyrosine) in strand β5 from the surface of the 455 face to the interior of the protein, thus propagating conformational changes from the active site to the 455 face (*1*). In Sma0114 Thr86, corresponding to the threonine in Y-T coupling, is too far from the active site to hydrogen bond to the phosphate and shows little structural difference between the inactive and activated states (Figure 4-4A). The conserved aromatic residue is replaced by Leu103, which switches from the inside to the outside of the Sma0114 structure on activation, the opposite of what is seen for the aromatic residue in classical Y-T coupling. Receiver domains that are substrates for HWE-kinases have a highly conserved PFxFATGY sequence motif that encompasses the threonine of Y-T coupling, Thr86 in Sma0114, and we proposed that it could be used as an alternative for communication between the active site and the 455 face (*31*). The NMR structure of the activated state, however, shows that residues of the PFxFATGY motif in strand β4 retain similar conformations to those seen in the inactive state, and are too far away from the active site to substitute for the Y-T coupling mechanism (Figure 4-4A). We can conclude that Y-T coupling is eliminated in Sma0114 and the conserved PFxFATGY motif appears not to be an alternative conduit of communication between the active site and output response face.

The only interaction between the 455 face and the active site that to our knowledge is conserved in all receiver domains, is that between Lys105 in loop $L_{\beta5\alpha5}$ and the phosphate or BeF_3^- ion. In the activated state of Sma0114, the repositioning of Lys105 to bind BeF_3^- is accompanied by a number of changes that affect the 455 face. These include a displacement of Leu103 in strand β5 from the interior to the surface of the protein, and movements of Phe107 and Glu111 towards the interior of the protein that bring helix α5 closer to the β-sheet core and to helix α1 in the activated structure. Although the role of Lys105 has traditionally been viewed as stabilizing the acyl phosphate (*4*), it also seems the most likely candidate to effectuate the communication between the active site and 455 face of Sma0114.

Because different receiver domains recognize a variety of downstream targets the structural changes accompanying activation are variable. In most receiver domains the largest changes induced by activation occur on the 455 face (Figure 4-3), however, this is not always the case. Activation of the bacterial sporulation factor Spo0F leaves the 455 face relatively unchanged. The largest structural changes occur in helix α1 (Figure 4-3), and these have been shown to play important roles in protein-protein interactions with the downstream targets of Spo0F (*39*). Activation of Sma0114 leads to structural changes in strand β5 and helix α5 of the 455 face (Figure 4-2), with the important difference that helix α4 of the typical receiver domain fold is replaced by a flexible loop. Additional changes extend to helix α1, which is also perturbed in the activated state of the three receiver domains NtrC, CheY, and Spo0F (Figure 4-3). We cannot identify if the rearrangement of helix α1 in Sma0114 occurs after Ca^{2+} binding or BeF_3^- binding,

although the perturbed conformation is maintained in the fully activated enzyme. Therefore, helix α1 may have a role in priming the enzyme for phosphorylation or participate in the output response by binding effectors in the fully activated state. Changes in secondary structure elements accompanying activation must lead to changes in the surfaces of receiver domains that enable them to bind their downstream target molecules. In the case of Sma0114, the sidechain rearrangements that accompany activation give the 455 face a more negatively charged character (Figure 4-5) that could be used to select positively charged target molecules.

Receiver domain dynamics on slow μs-ms timescales have been investigated using relaxation dispersion methods (*6, 20, 58*). These studies showed that the inactive and activated conformations exist in a dynamic equilibrium, that is shifted in favor of the activated state when the active site binds ligands (*6*). Our results for Sma0114 are consistent with this two-state allosteric mechanism of activation. The apo (Figure 4-8A) and Ca^{2+}-bound enzyme (Figure 4-8B) exhibit $R2_{ex}$ line-broadening contributions for NMR signals from residues in the active site and 455 face. The $R2_{ex}$ contributions are suppressed in the Sma0114•Ca^{2+}•BeF_3^- ternary complex, where the population of the activated state dominates (Figure 4-8C).

By contrast, there have been few studies of receiver domain dynamics on the fast ps-ns timescales accessible by ^{15}N relaxation measurements. Spo0F was characterized only in the apo-state where dynamics on the ps-ns timescale were shown to be constant except at the chain termini (*19*). For NtrC, ps-ns dynamics were characterized for both the inactive and activated forms of the enzyme, and were shown to be constant except at

the chain termini (*6*). In the case of Sma0114 we characterized ps-ns dynamics for three forms of the protein: the apo-state, the Ca^{2+}-bound state, and the activated state (Figure 4-6). Both the Sma0114•Ca^{2+} binary complex (Figure 4-6B) and the Sma0114•Ca^{2+}•BeF_3^- ternary complex (Figure 4-6C) show an increase in the number of residues exhibiting flexibility on the ps-ns timescale compared to the apo-protein (Figure 4-6A). These results suggest an entropy-compensated binding mechanism for the BeF_3^- phosphate analog, where the decrease in entropy due to ligand binding in the active site is compensated by increased dynamics in other parts of the structure. Entropy-compensated binding was first hypothesized in 1963 for insulin dimerization (*53*). The mechanism has since been supported by NMR relaxation studies showing increased dynamics accompanying ligand binding in a number of proteins (*54, 55, 59*), including the enzymes MutT (*56*), stromelysin (*60*), and topoisomerase I (*61*).

In the case of Sma0114, entropy-compensated binding is likely to be exploited for a functional role, as most of the residues that show increased dynamics in the activated state are part of the 455 face of the protein. These include residues Thr104-Phe107 in strand β5 and loop $L_{\beta5\alpha5}$ and Val115-Val117 in helix α5. The mobile loop, (Gly89-Leu93) that replaces helix α4, together with the above segments give the 455 face in Sma0114 a flexible character (Figure 4-7B). Flexibility in the 455 face could be used to bind downstream effector proteins through an induced-fit mechanism, initially proposed by Dyson and Wright (*62, 63*). In an induced-fit mechanism, weak interactions with diffuse specificity are first formed using the dynamic segments of a binding surface and strengthen as the appropriate ligand is drawn closer to the receptor and the complement

of correct interactions are formed. The dynamically disordered surface in Sma0114 may facilitate binding a range of molecules each having a common molecular shape recognized by the 455 face.

Chapter 5. Materials and Methods

Sma0114 Expression and Purification.

The *sma0114* gene was amplified from *S. meliloti* strain Rm1021, cloned into the pET28(+) plasmid, and transformed into *Escherichia coli* strain BL21(DE3)pLysS. Cells were grown in 1 liter of M9 medium containing 1.0 g ^{15}N-NH_4Cl and/or 3 g of ^{13}C-glucose supplemented with kanamycin to a final concentration of 25 μg/ml. Following incubation at 30°C to mid-log phase, the recombinant His-tagged Sma0114 protein was induced by the addition of isopropyl β-D-1 thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM. Cells were harvested 18 hours after induction and lysed by sonication in affinity column binding buffer (100 mM Tris-HCl, 0.5 M NaCl, pH 7.4) containing lysozyme and 1 mM phenylmethanesulfonyl (PMSF). The cell debris was sedimented and the His_6-Sma0114 fusion protein in the supernatant was purified by nickel affinity chromatography on a 1 ml HisTrap column. Following elution, fractions containing His_6-Sma0114 were pooled and cleaved overnight at 23°C with 10U/ml of bovine thrombin. PMSF was subsequently added to inactivate thrombin, and the Sma0114 component was further purified by gel filtration on a Superdex S75 column. The fractions eluted from this column were run over a 1 ml HisTrap column to remove any uncut His_6-Sma0114. The remaining flow-through containing Sma0114 at > 95% purity by SDS-PAGE, was dialyzed extensively against the buffer used for NMR (50 mM sodium phosphate pH 6.0). Following dialysis, samples were concentrated for NMR using 10,000 molecular weight cut-off centrifugal filter units.

The Sma0114 construct used for these studies contains the extra three N-terminal residues Gly-Ser-His from the thrombin cleavage site used to purify the protein. These

residues are included in our numbering scheme for the protein, which runs from residues Gly1 to Val123 with only residues Met4-Val123 part of the wild type Sma0114 amino acid sequence. We did not detect NMR signals from the amide protons of the extra residues Gly1-His3, suggesting they are disordered.

NMR Sample Preparation

NMR samples of the inactive enzyme were dissolved in 50 mM sodium phosphate pH 6.0 supplemented with 0.02% NaN_3 to prevent bacterial growth and 1 mM dithiothreitol (DTT) to prevent disulfide formation due to the sole cysteine at position 29 in the protein. For NMR spectroscopy on the activated enzyme, ^{15}N-Sma0114 and ^{15}N,^{13}C-Sma0114 samples were dissolved in 50 mM 2-(N-morpholino)ethanesulfonic acid (MES), buffered to pH 6.0. MES buffer was used for experiments on the binary Sma0114•Ca^{2+} and ternary Sma0114•Ca^{2+}•BeF_3^- complexes rather than phosphate, to prevent precipitation of Ca^{2+}. Samples had 500 μM concentrations of protein. All NMR experiments were done at a temperature of 37°C. NMR metal titrations were performed using ^{15}N-labeled samples of Sma0114, over a range of concentrations between 0 and 500 mM $MgCl_2$. Due to the high 150 mM concentration of $MgCl_2$ needed to saturate the enzyme, we also looked at the binding of $CaCl_2$ over a concentration range between 0 and 10 mM. With $CaCl_2$, saturation of Sma0114 was achieved at a metal ion concentration of 1.5 mM.

The Sma0114•Ca^{2+}•BeF_3^- ternary complex was prepared using the approach of Yan, et al. (*49*). BeF_3^-, generated in situ from the addition of NaF to $BeCl_2$, mimics the acyl phosphate linkages formed by the active site aspartates of receiver domains (*11, 39,*

49, 50). Starting from samples of Sma0114 containing 1.5 mM $CaCl_2$, we added NaF in increments of 5 mM and observed no changes in ^{1}H-^{15}N HSQC spectra of the protein up to a final concentration of 30 mM NaF. Thus the presence of NaF, which is needed to generate BeF_3^- from $BeCl_2$, does not perturb the structure of Sma0114. We next added $BeCl_2$ in 1 mM increments and found that a 5 mM concentration was sufficient to saturate the active site, as monitored by the point at which no further chemical shift changes were seen in ^{1}H-^{15}N HSQC spectra. Thus the final sample conditions for the Sma0114•Ca^{2+}•BeF_3^- ternary complex were 500 μM Sma0114, 1.5 mM $CaCl_2$, 30 mM NaF, and 5 mM $BeCl_2$.

Inactive Sma0114 Nuclear Magnetic Resonance Spectroscopy

Samples of 250 μl volumes were taken up Shigemi microcells. Spectra were recorded on a Varian INOVA 600 MHz spectrometer equipped with a cryogenic probe. The temperature for all experiments was 37°C. NMR assignments were obtained using 2D and 3D experiments (*64*) implemented in the Varian Protein Pack. Starting from the 2D ^{1}H-^{15}N HSQC spectrum, backbone ^{1}H, ^{15}N and ^{13}C resonances were assigned using two redundant sequential walk pathways. The first used 3D HNCACB, HNCA and HN(CO)CA data to establish Cα and Cβ connections across peptide bonds. The second pathway used HNCO and HN(CA)CO experiments to connect C' carbons across peptide bonds. Sidechain assignments were obtained from 3D HCCH-TOCSY, H(CCO)NH, C(CO)NH, HNHA, HNHB, ^{1}H-^{15}N TOCSY-HSQC, and ^{1}H-^{13}C HSQC-NOESY data. Stereo-specific assignments for methylene protons were based on short mixing time (25 ms) 2D NOESY, 3D ^{1}H-^{15}N NOESY-HSQC and HNHB experiments as described (*65*).

Aromatic ^{1}H and ^{13}C resonances were obtained from 2D ^{1}H-^{13}C HSQC, DFQ-COSY and TOCSY data. 2D NOESY experiments (150 ms mixing time) using a sample in D_2O were used to connect aromatic resonances with previously assigned aliphatic resonances. Sidechain amide resonances were linked through intra-residue NOEs to previously assigned aliphatic protons using 3D ^{1}H-^{15}N NOESY-HSQC data. DSS was used as an internal reference for ^{1}H resonances, and ^{13}C and ^{15}N resonances were referenced indirectly (*66*). Spectra were processed and analyzed using FELIX-NMR and iNMR. The limits of secondary structure elements were determined using $^{3}J_{HNH_\alpha}$ - coupling constants (*67*), TALOS chemical shift analysis (*29*) and hydrogen-deuterium exchange data.

Inactive Sma0114 Chemical Shift Assignments

^{1}HN, ^{15}N, $^{13}C'$, Cα, Cβ, Hα and Hβ resonances were assigned for 94% of the amino acids in Sma0114. In total, we obtained assignments for 84% of all carbons (90% of aliphatic carbons and 65% of aromatic carbons), 84% of all protons (84% aliphatic and 85% of aromatic protons) and 75% of all nitrogens. Four of 118 non-proline residues were not observed in the ^{1}H-^{15}N HSQC spectrum: Leu108, and residues Gly1-His3 which are not part of the wild type sequence but were introduced as part of a thrombin cleavage site used in purification of Sma0114. Chemical shift assignments are deposited in the Biological Magnetic Resonance Bank (http://www.bmrb.wisc.edu) under accession code BMRB-16905.

Chemical shifts for the Activated Enzyme

The 2D ^{1}H-^{15}N HSQC of the inactive enzyme (Figure 2-2) was used as a reference for assigning resonances in the ternary Sma0114•Ca^{2+}•BeF_3^- complex. Having complete assignments for the inactive protein (Chapter 2) facilitated NMR assignments for the active form since only a subset of resonances undergo changes upon activation. Figure 5-1 shows chemical shift changes in ^{1}H-^{15}N HSQC spectra for the Ca^{2+}-bound enzyme (A) and the fully activated (Ca^{2+} and BeF_3^- bound) enzyme (B).

Changes in sidechain chemical shifts for the activated enzyme were determined from ^{13}C-HCCH-TOCSY spectra which correlates all side-chain protons with alpha or beta protons of individual amino acids while allowing the assignment of sidechain carbons (*68*). Chemical shift assignments for the Sma0114•Ca^{2+}•BeF_3^- ternary complex have been deposited in the BMRB database under accession code 19286.

NMR Structure Determination for Inactive Sma0114

3D ^{15}N- and ^{13}C-editted NOESY experiments (*64*) were used to obtain NOE-based distance restraints. Long-range HNCO (*69*) and deuterium isotope exchange experiments were used to identify hydrogen bonds. Restraints for the backbone dihedral angles φ and ψ were calculated from the assigned HN, Hα, N, Cα, Cβ and C' chemical shifts using the program TALOS (*29*). Additional 3D HNHA data were collected to check φ dihedral angles. Side-chain χ1 dihedral angles and stereospecific assignments for prochiral methylene protons were determined from 3D HNHB data (*64*) and short-mixing time 2D NOESY spectra (*65*). Stereospecific assignments for the prochiral methyl groups of Leu and Val residues were obtained from a sample fractionally labeled with

10% ^{13}C-glucose (*70*). NMR spectra were collected on a Varian Inova 600 MHz instrument equipped with a cryogenic probe. The NMR structure of Sma0114 was calculated using the program X-PLOR (v. 3.851) (*34*) based on 1627 experimental restraints (Table 3-1). The 20 lowest energy structures have been deposited in the PDB under accession code 2LPM.

NMR Structure Determination for BeF_3^- activated Sma0114

Chemical shifts for the Sma0114•Ca^{2+} binary complex were obtained from the titration of ^{1}H-^{15}N crosspeaks in 2D ^{1}H-^{15}N HSQC spectra of Sma0114 as a function of $CaCl_2$ concentration (*31*). NMR assignments for the Sma0114•Ca^{2+}•BeF_3^- ternary complex were made from 2D and 3D NMR experiments as previously described for the apo protein. Distance restraints for structure calculations were obtained from 3D ^{15}N- and ^{13}C-edited NOESY experiments (*64*) collected on a Varian Inova 800 MHz instrument, and a 2D ^{1}H-NOESY experiment for the aromatic protons recorded at 600 MHz. The mixing time for all NOESY experiments was 150 ms. Hydrogen bonds were identified from deuterium isotope exchange experiments recorded on the ternary complex dissolved in 99.96% D_2O. Hydrogen bond donors were identified as those residues seen in ^{1}H-^{15}N HSQC spectra after deuterium isotope exchange. The protection pattern for the activated state was compared to that seen in the inactive enzyme and no changes were identified. We therefore determined that the hydrogen bond donors and acceptors did not change due to activation. Backbone φ and ψ dihedral angles were calculated from the assigned HN, Hα, N, Cα, Cβ and C' chemical shifts using the program DANGLE in CCPNMR-analysis (*71*).

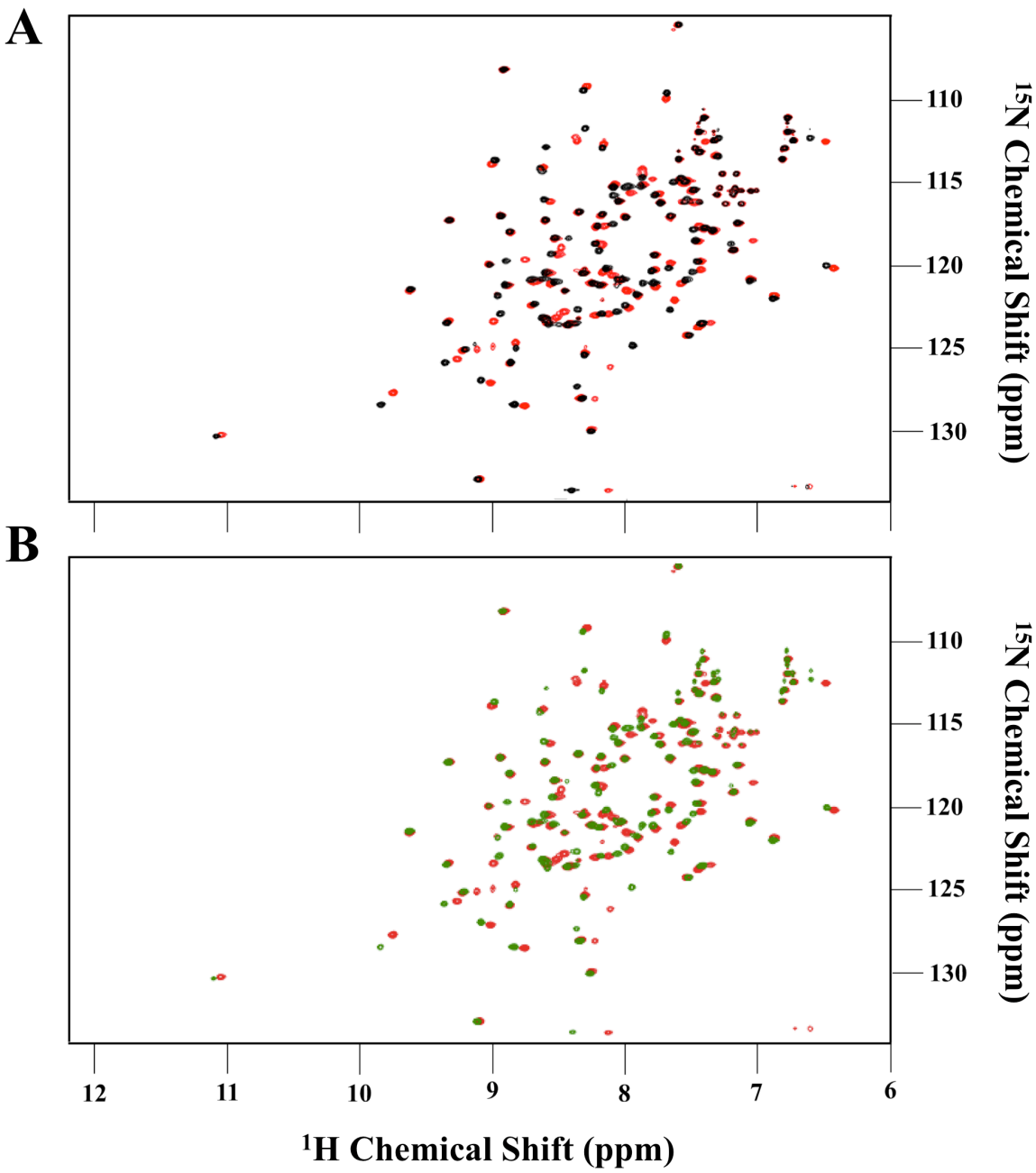

Figure 5-1 (**A**) ^{1}H-^{15}N HSQC of the free protein (black) and the Ca^{2+} - bound protein (red). (**B**) ^{1}H-^{15}N HSQC of the Ca^{2+} - bound (red) and the BeF_3^- activated protein (green).

The NMR structure of the activated Sma0114•Ca^{2+}•BeF_3^- ternary complex, was calculated using the program X-PLOR (v. 3.851) (*34*) based on 1791 experimental restraints (Table 4-1). Calculations were performed using Ca^{2+}, Be, and F atomic radii from the X-PLOR library (*34*) to include the divalent metal ion and BeF_3^-. Because there are no literature examples available for a bond between Ca^{2+} and BeF_3^-, we set the distance restraint between the two atoms at 2.35 Å, equal to the sum of the ionic radii for Ca^{2+} (0.99 Å) and F^- (1.36 Å). The distance restraint was given an uncertainty bound of 0.1 Å and the three fluorine atoms in BeF_3^- were treated as ambiguous. We next included an ambiguous restraint of 1.5 ± 0.3 Å between the two sidechain carboxylic acid oxygens of Asp60 (OD1 or OD2) and the beryllium atom. The restraint was based on distances between the beryllium atom and the closest oxygen of the equivalent aspartate in representative X-ray structures of receiver domains (PDB codes: 1FWQ, 2A9O, 2PL1, 3NNN). Bond distances in these structures ranged from 1.5 to 2.0 Å, with an average distance of 1.7 Å. We used the same approach to restrain the three possible Ca^{2+} ligands in the protein, Glu15-Asp16-Glu17. An analogous stretch of two, or three acidic residues in a row is found in the receiver domain structures listed above. To determine which of the three acidic residues in Sma0114 could be ligands for Ca^{2+}, we carried out structure calculations with the distance restraint set for the complete structure but excluding restraints to Ca^{2+} or BeF_3^-. Superimposition of the two structures showed that both Glu15 and Asp16 were within bonding distance of the Ca^{2+} atom in the structures calculated without restraints to the ligands. These two residues are also within bonding distance to the Ca^{2+} atom in the apo enzyme, determined from a comparison of the NMR structures

of inactive and activated states (*31*). That is, the two residues already appear to be positioned to bind the metal in the inactive apo state. Glu17 points away from the Ca^{2+} atom but the conformation of this residue is poorly defined due to R2 line broadening. We therefore decided to also include Glu17 as a potential Ca^{2+} binding ligand. Distance restraints were set to 2.45 ± 0.15 Å based on literature values of typical bonds between carboxylic acid groups and Ca^{2+} in known protein structures (*72*), which range between 2.3-2.6 Å. The distance restraints where treated as ambiguous with respect to the oxygens bound to the Ca^{2+} atom: Glu15 (OE1 or OE2), Asp16 (OD1 or OD2), and Glu17 (OE1 or OE2). Finally, we introduced a restraint of 2.5 ± 0.5 Å between the NZ atom of Lys105 and the fluorine atoms of BeF_3^-, which were treated as ambiguous in the calculations. An analogous lysine is found bound to BeF_3^- in all BeF_3^- activated receiver domain structures, with a bonding distance between 2.8 and 3.0 Å. To test the effect of this restraint we calculated a set of structures in which only the Lys105-BeF_3^- restraint was removed. Without the restraint, the lysine sidechain was positioned in the interior of the structure but the distance of 6.0 Å to BeF_3^- precluded a bonding interaction. Given that a buried charged lysine would be highly energetically unfavorable (*73*) as there are no other nearby negatively charged groups that could complement the positive charge on Lys105, we felt the restraint to the BeF_3^- ion was justified. The restraints from the protein to the Ca^{2+} and BeF_3^- ligands described above, introduced no violations in the structure calculations. The RMSD between the structures calculated with and without restraints to the ligands was 0.80 Å, comparable to the 0.88 Å value for the spread in the ensemble of

activated state NMR structures (Table 4-1). This indicates that the derived restraints to the ligands do not affect the overall structure, and that the structure with the restraints to the ligands is compatible with the experimental NOE, hydrogen-bond, and dihedral restraints. The 26 lowest energy structures for the activated state of Sma0114 have been deposited in the PDB under accession code 2M98.

NMR Relaxation Measurements

Backbone dynamics of Sma0114 were investigated using ^{15}N R1, R2 and ^{1}H-^{15}N NOE experiments. Longitudinal relaxation rates (R1) were characterized using relaxation delays of 0.02, 0.05, 0.13, 0.21, 0.31, 0.5, 0.71 and 1.0 s. Transverse relaxation rates (R2) were measured using relaxation delays of 0.01, 0.03, 0.05, 0.07, 0.09, 0.15, 0.25 and 0.35 s. ^{1}H-^{15}N NOE values were determined from experiments in which the proton signals were saturated (s) for 4 s and control experiments (c) in which the saturation period was replaced by an equivalent preacquisition delay. Spectra were acquired in an interleaved manner on a Varian Inova 600 MHz instrument for all three relaxation data sets. Relaxation rates were calculated from least-squares fits of the data to an exponential decay model (Eq. 1) where I is the intensity for the relaxation period τ, I_0 is the initial amplitude and $R_{1,2}$ corresponds to the relaxation rate R1 or R2.

$$I = I_0 \bullet \exp(-\tau/R_{1,2}) \qquad \text{Equation 1}$$

Experimental uncertainties in relaxation parameters were taken as the standard errors of the fits. ^{1}H-^{15}N NOE values were calculated according to Equation 2, where I(s) is the

crosspeak intensity in the experiment with saturation (s) and I(c) is the crosspeak intensity without saturation (c).

$$NOE = I(s)/I(c) \qquad \text{Equation 2}$$

The errors for the ^{1}H-^{15}N NOE experiment were determined as described previously (*57, 74*). R1, R2 and ^{1}H-^{15}N NOE values were used as input for Model-Free calculations (*42*) using the program *Tensor2 (32)*.

Bioinformatics Analysis of Sma0114

Analysis of 273 HWE-kinases collected from the MiST 1 database (*75*) showed that 88 of the kinases had associated receiver domains. Receiver domains were considered to be associated with an HWE-kinase if the genes were fused or immediately upstream or downstream of the kinase. The cognate receiver domains of the 88 HWE-kinases identified in this way, were analyzed using the secondary structure prediction program PsiPred (v 2.6) (*43*). The final alignment excluded 5 of the 88 sequences because they were less than 100 amino acids in length and did not appear to encode full-length receiver domains. As a control, 8 receiver domain sequences of known structure were employed in the alignment to verify that the α_5/β_5 motif could be detected accurately by PsiPred (*43*).

Chapter 6. Future Directions

Identifying the downstream target that Sma0114 interacts with remains a key objective. Complementary genetic and bioinformatics work on the downstream function of Sma0114 is underway. A deletion mutant removing the gene for Sma0114 (Δ*sma0114*) led to increased gene transcription in *S. meliloti* whereas deleting the Sma0113 HWE-kinase (Δ*sma0113*) and a double mutant (Δ*sma0113sma0114*) led to no change in gene transcription. Taken together these data show that Sma0114 requires functional Sma0113, its cognate histidine kinase, in order to exert its phenotype. Current studies building off of these results suggest that the primary role of Sma0114 is to inhibit Sma0113. The working theory is that phosphorylation of Sma0113 induces the phosphorylation of Sma0114 and leads to dissociation of the two proteins. Once Sma0114 is dissociated from Sma0113, Sma0113 is able to activate another receiver domain protein. Genetic analysis should be performed to identify other receiver domain partners that may be involved in a phospho-relay with Sma0113 once the kinase is phosphorylated.

In light of the fact that the inhibition of Sma0113 may be the primary role of Sma0114, further studies on the structure of Sma0113 should be performed. A necessary first step is to make Sma0113 soluble because its insolubility has precluded analysis of the enzyme's structure and function. The role of the His, Trp and Glu residues important in the HWE-kinase subfamily should also be explored in Sma0113. Mutants for these residues should be designed. In this way parallel mutant experiments can be compared to those done for the HWE-kinase BphP2 (*18*) to determine the importance of the three

residues in phospho-transfer. In BphP2 point mutations for each HWE residue eliminated autophosphorylation (*18*), so that we would expect to see a similar profile in Sma0113. Structural analysis of Sma0113 would enable an in-depth study of the role of each residue in the 3-dimensional form of the enzyme. Comparison of HWE-kinases to the histidine kinase superfamily would also be accomplished by solving the structure of Sma0113. Furthermore, titrating Sma0113 into ^{15}N-labelled Sma0114 and recording changes in the ^{1}H-^{15}N HSQC of Sma0114 can help define the interface between the two enzymes.

In order to explore the role of Lys105 in Sma0114 as the conduit of communication between the active site and 455 face, the residue should be mutated. A Lys105Arg mutant can be designed to maintain a positive charge at the position to test if any basic residue will function as a conduit. A Lys105Leu mutant would allow us to further test the importance of the positive charge at that position. In these mutants the ability to bind BeF_3^- as well as conformational changes in the 455 face should be determined. ^{15}N relaxation experiments can be used to explore ps-ns dynamics and compare S^2 order parameters of wildtype Sma0114 and Lys105 mutants.

The role of residues in the PFxFATGY motif should also be explored via mutagenesis. The Thr, Gly and Tyr residues that form the C-terminal end of the motif and form the beginning of $L_{\beta4\beta5}$ should be explored as they precede the dynamic loop replacing helix α4 in Sma0114. Point mutations for each residue should be tested for the ability to bind BeF_3^- and any changes in Sma0113 inhibition. Further NMR studies including a ^{1}H-^{15}N HSQC and ^{15}N relaxation should be performed to identify changes in conformation and dynamics as a result of the mutations.

Reprinted from Sheftic et al. (*76*)
I performed kinetic assays and electron microscopy (Figures A1-1, A1-2, A1-3)
Jessica Snell performed NMR experiments (Figure A1-4)
Dr. Suman Jha performed ITC experiments (FigureA1-5)

Appendix One. Misfolding of SEVI Inhibited by Physiological Metal Ions

Abstract

SEVI (semen derived enhancer of viral infectivity), a naturally occurring peptide fragment of prostatic acid phosphatase, enhances HIV infectivity by forming cationic amyloid fibrils that aid the fusion of negatively charged virions and target cell membranes. Cu(II) and Zn(II) inhibit fibrillization of SEVI in a kinetic assay employing the fibril-specific dye Thioflavin T. Transmission electron microscopy (TEM) suggests that the metals do not affect fibril morphology. NMR shows that the metals bind to histidines 3 and 23 in the SEVI sequence. Isothermal titration calorimetry (ITC) experiments indicate that SEVI forms oligomeric complexes with the metals. Dissociation constants are micromolar for Cu(II) and millimolar for Zn(II). Because the Cu(II) and Zn(II) concentrations that inhibit fibrillization are comparable to those found in seminal fluid the metals may modulate SEVI fibrillization under physiological conditions.

Introduction

SEVI is a naturally occurring 39-residue (4.6 KDa) proteolytic fragment of human prostatic acidic phosphatase ($PAP_{248-286}$) that was discovered in a screen of semen compounds for effects on HIV infectivity (*77*). The SEVI peptide is present at a concentration of ~35 μg/ml in semen, and can assemble into amyloid fibrils with a cross β-sheet structure at concentrations above 2 μg/ml (*77*). The fibrils attach to virions and facilitate their fusion to target cells, thereby increasing HIV infectivity up to 400,000-fold (*77*). As such, the SEVI peptide could be an important facilitator of HIV transmission.

The positively charged amyloid fibrils formed by SEVI are thought to promote viral fusion by reducing electrostatic repulsion between negatively charged virus and target cell membranes (*77-79*). The activity of SEVI appears to be general, in that it promotes cell-fusion of a variety of HIV variants (*77*) as well as human XMRV, a retrovirus that may play a role in prostate cancer (*80*). SEVI also boosts infectivity nonspecifically for a number of other retroviruses - a finding that may have applications to therapeutic gene transfer (*81*). The generality of the charge-screening mechanism is further emphasized by the observation that designed peptides, with sequences unrelated to SEVI, retain the ability to enhance HIV infectivity as long as they assemble into fibrils with cationic surfaces (*82*).

Semen contains a very complex mixture of compounds (*77, 83-85*). It has been shown recently that anionic buffer molecules in human seminal plasma accelerate fibrillization (*86*), whereas other components such as proteases inhibit fibrillization by degrading SEVI (*84*). In this work, we focus on how SEVI fibrillization is affected by heavy metals since these are found at large concentrations in seminal plasma and have

been shown to modulate the fibrillization of other amyloidogenic proteins and peptides including Aβ (*87*), amylin (*88, 89*), α-synuclein (*90*), β-microglobulin (*91*), insulin (*92*), and prion protein (*93*).

Results

Cu^{2+} and Zn^{2+} Inhibit SEVI Fibrillization

Figure A1-1 shows the effects of increasing $CuCl_2$ (Figure A1-1A) and $ZnCl_2$ (Figure A1-1B) concentrations on SEVI fibrillization. Both metals inhibited fibrillization compared to control samples of SEVI with a 5 mM concentration of the metal chelator EDTA. Additional control experiments were used to establish that the presence of EDTA had no effect on the fibrillization kinetics. The Zn^{2+} and Cu^{2+} ions are good ligands for histidines (*88*), which are present at positions 3 and 23 in the SEVI amino acid sequence. In addition to these metals, the effects of 1 mM Ca^{2+} and Mg^{2+} which are poor ligands for histidine imidazole rings (*88*) were explored. Inhibition of fibrillization was observed with 1 mM $MgCl_2$ although to a lesser extent than with $CuCl_2$ or $ZnCl_2$. By contrast, $CaCl_2$ had negligible effects compared to the control samples without metals.

Nucleation of a fibrillization reaction is entropically disfavored due to the loss of rotational and translational degrees of freedom when monomers aggregate. The lag phase of the fibrillization curve represents the period of time before nucleation occurs. Once a critical mass or nucleus forms, the reaction enters a growth phase in which addition of monomers becomes enthalpically favorable as new non-covalent interactions form during the fibril elongation process. The reaction ultimately reaches a steady state plateau determined by the solubility of the monomers. The effects of Zn^{2+} and Cu^{2+} on the kinetic parameters characterizing SEVI fibrillization reactions (lag times, elongation rates, and plateaus) are shown in Figure A1-2. The most striking effect is seen on the lag times for fibrillization (Figure A1-2A). These increase 4 to 8-fold in the presence of the metals

compared to control samples containing SEVI alone, suggesting that the nucleation part of the reaction is inhibited in the presence of metals.

The lag times show a constant 4-fold increase with Cu^{2+} concentrations in the range between 0.001 – 0.1 mM, consistent with a low dissociation constant for Cu^{2+} (see ITC section below). At the largest concentration of 1 mM $CuCl_2$, there is a reversion to shorter lag times but these nevertheless remain about 2-fold longer than for SEVI without metals. Lag times also increase with increasing Zn^{2+} concentrations, reaching a plateau at 0.1-1.0 mM $ZnCl_2$. The data are consistent with a Zn^{2+} dissociation constant in the range of ~1 mM, a value some 3 orders of magnitude larger than that for Cu^{2+}.

At the lowest 0.001 mM concentrations of Cu^{2+} and Zn^{2+} tested, there is a two-fold reduction in fibril elongation rates compared to SEVI alone. This suggests that in addition to interfering with nucleation the metals hinder the addition of SEVI monomers to growing fibrils. As the metal concentration is increased up to 1 mM, however, the elongation rates show no further decrease, within experimental uncertainty (Figure A1-2B). When the fibrillization reactions enter a steady state, the fluorescence plateaus show a slight decrease with increasing Cu^{2+} concentrations, whereas they increase with increasing Zn^{2+} concentrations (Figure A1-2C). Experiments done with a lower SEVI peptide concentration of 1 mg/ml showed the same increase in lag times and decrease in rates but the fluorescence plateaus were decreased for both Cu^{2+} and Zn^{2+}. Fluorescence plateaus often correlate with the amount of fibrils formed but can also be influenced by differences in fibril morphology that affect interactions with the dye ThT.

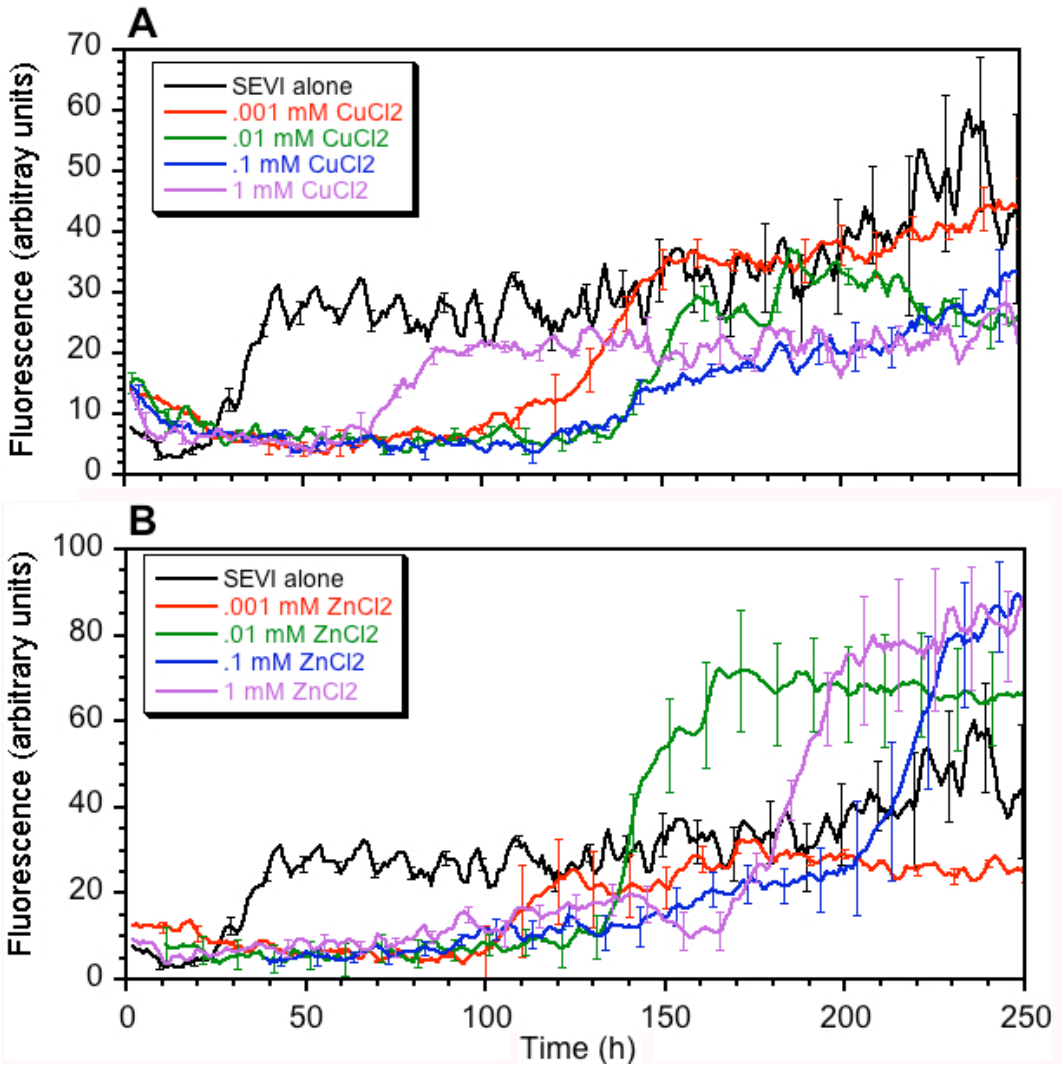

Figure A1-1 Inhibition of SEVI fibrillization with increasing concentrations of (**A**) $CuCl_2$ and (**B**) $ZnCl_2$. All experiments were done with 2 mg/ml SEVI in PBS buffer, pH 7.4 at a temperature of 37°C. Error bars are the standard errors of the mean values calculated from triplicate runs and are shown every tenth point for clarity.

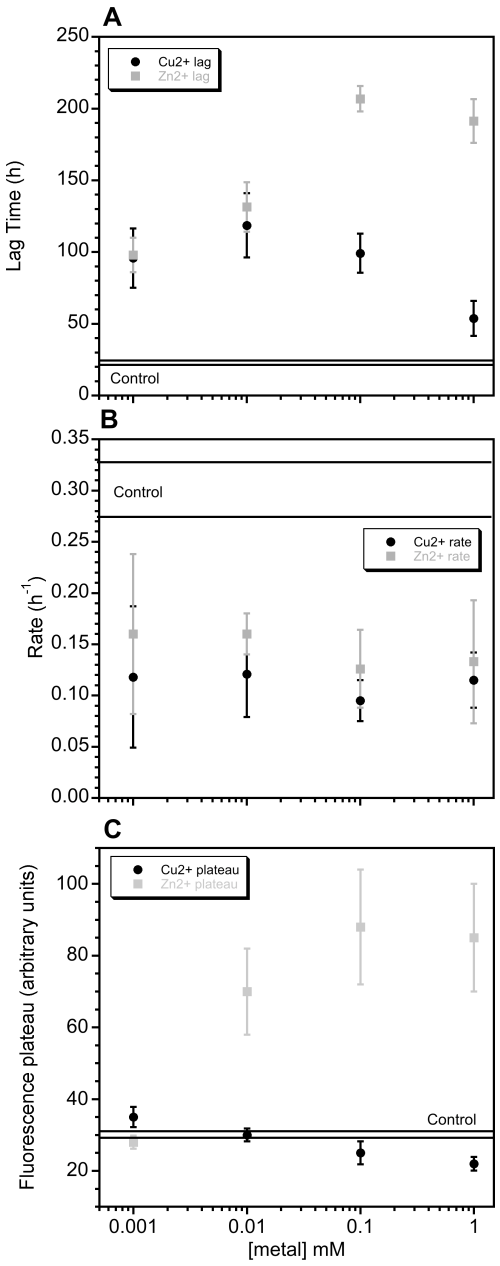

Figure A1-2 Semi-logarithmic plots showing the $CuCl_2$ (black) and $ZnCl_2$ (gray) concentration dependence of kinetic parameters for SEVI fibrillization derived from the data in Figure A1-1. (**A**) Lag times. (**B**) Fibril growth rates. (**C**) ThT fluorescence plateaus. Uncertainty bars represent the standard errors from kinetic parameters of experiments run in triplicate. The horizontal lines marked "control" are ranges for parameters obtained from SEVI experiments without metals.

Even if the larger fluorescence plateaus in the presence of 0.01-1.0 mM Zn^{2+} are due to larger numbers of fibrils in the presence of large Zn^{2+} concentrations, the lag times under these conditions are increased eight to nine fold (Figure A1-2A) so that Zn^{2+} would be a potent inhibitor of fibrillization on physiological timescales.

Fibril Morphology is Retained but the Amount of Fibrils is Lowered in the Presence of Zn^{2+} and Cu^{2+}

Having established that Zn^{2+} and Cu^{2+} inhibit fibril formation, we next wanted to see if the metals affected the morphologies of the fibrils. Figure A1-3 compares control SEVI fibrils formed in the absence of metals (Figure A1-3A) to fibrils formed in the presence of 1 mM $ZnCl_2$ (Figure A1-3B) or $CuCl_2$ (Figure A1-3C). High magnification TEM images show that the metals have no apparent effects on fibril morphology (Figure A1-3). Lower magnification images given in the insets for Figure A1-3 suggest that the amount of SEVI fibrils is reduced in the presence of 1 mM $ZnCl_2$ or $CuCl_2$ compared to the control.

NMR Mapping Identifies the Two Histidines in SEVI as the Ligands for Zn^{2+} and Cu^{2+}

Previous NMR studies reported that the SEVI peptide is intrinsically unfolded at 37°C but folds into an α-helical structure in 50% trifluoroethanol similar to that of the corresponding 248-286 segment in the parent protein human prostatic acid phosphatase (*94*). The changes in the NMR spectrum accompanying binding of diamagnetic zinc to SEVI are manifested as line-broadening, caused by intermediate exchange between the free and bound forms of the peptide. To better characterize the binding site, 2D ^{1}H-^{15}N

HSQC spectra of a 3 mM SEVI sample at natural isotopic abundance were acquired. Spectra were recorded in the absence of metals, in the presence of a 2:1 peptide:Zn^{2+} ratio, and at a 1:1.3 peptide:Zn^{2+} ratio (Figure A1-4A). Losses in ^{1}H-^{15}N HSQC crosspeak volumes due to intermediate exchange between the free and bound peptide are localized to residues 2-6 and 23-24, flanking the two histidines at positions 3 and 23.

Line broadening is observed at substoichiometric concentrations of Zn^{2+}, and there are only small additional decreases of crosspeak volumes as the metal concentration becomes saturating. The magnitude of the line broadening effects is similar for the regions surrounding His3 and His23, suggesting the two histidines bind Zn^{2+} simultaneously (Figure A1-4B). Consistent results implicating His3 and His23 as the ligands for Zn^{2+} were seen in ^{1}H-^{13}C HSQC spectra of SEVI, except that less residues could be analyzed due to the poorer chemical shift dispersion in the carbon spectrum. We obtained analogous results for the binding of paramagnetic Cu^{2+} to SEVI except that a larger number of crosspeaks were broadened in the ^{1}H-^{15}N HSQC and ^{1}H-^{13}C HSQC spectra. Although the sequence profiles are less well-defined than with Zn^{2+}, the NMR data show that His3 and His23 also serve as the ligands for Cu^{2+}.

ITC Shows SEVI Forms Oligomeric Complexes with Cu^{2+} and Zn^{2+}

ITC experiments were conducted to obtain information on the stoichiometry and thermodynamics of metal binding (Figure A1-5). The titration for Cu^{2+} is well-defined and indicates that the metal binds with a substoichiometric Cu^{2+}/SEVI ratio of $n = 0.3$ and a dissociation constant of 2 μM (Figure A1-5A).

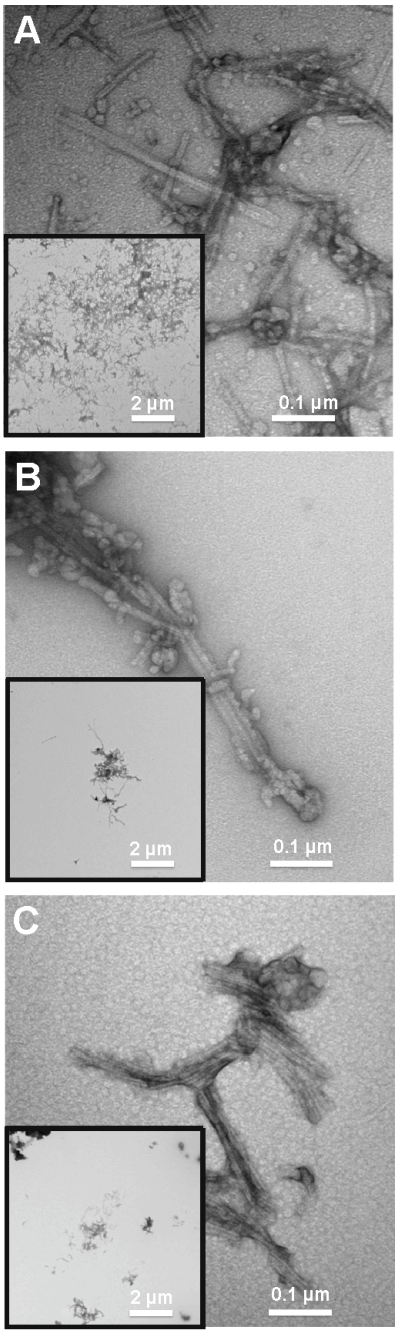

Figure A1-3 TEM images of SEVI fibrils. Images at 150,000x magnification. Control experiment showing SEVI fibrils grown in the absence of metals (**A**), in the presence of 1 mM $ZnCl_2$ (**B**), and in the presence of 1 mM $CuCl_2$ (**C**). The insets for each panel show images obtained at a lower magnification of 11,000x.

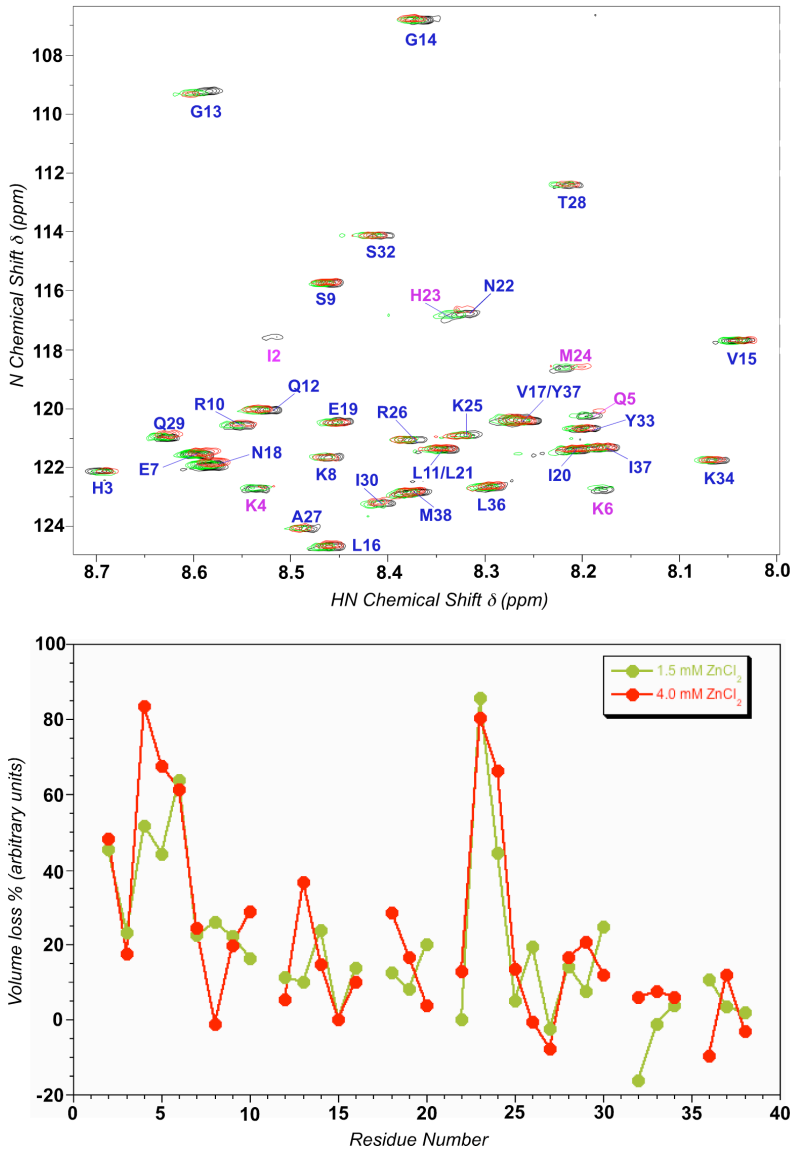

Figure A1-4 Effects of $ZnCl_2$ on the NMR spectrum of SEVI. (**A**) The three superimposed natural ^{15}N abundance ^{1}H-^{15}N HSQC spectra in the absence of zinc (black contours), 1.5 mM $ZnCl_2$ (green contours) and 4 mM $ZnCl_2$ (red contours). NMR assignments for SEVI in blue text indicate peaks that are unaffected and in purple text for residues that experience NMR signal broadening in the presence of zinc. (**B**) Plot of crosspeak volume loss (%) with $ZnCl_2$, as a function of residue position in the SEVI sequence.

The value of $n = 0.3$ suggest SEVI binds Cu^{2+} as a trimer, in which His3 and His23 from three SEVI monomers constitute the ligands of an octahedral metal coordination site. ΔG and ΔS values for the Cu^{2+} binding reaction were -32.5 ± 1.6 kJ/mol and 39.9 ± 5.4 J/(mol•K) respectively.

The ITC data indicates that SEVI binds Zn^{2+} three orders of magnitude weaker than Cu^{2+}. We obtained an apparent dissociation constant of 1.7 ± 0.9 mM, however, this value is subject to considerable uncertainty due to weak affinity of SEVI for Zn^{2+}. ΔG and ΔS values for the Zn^{2+} binding reaction were -15.8 ± 8.4 kJ/mol and -163 ± 234 J/(mol•K) respectively. Similarly, the *n*-value obtained from ITC of 0.45 ± 0.40 indicates substoichiometric binding but does not allow us to establish precisely how many SEVI molecules are bound to a Zn^{2+} ion. Whereas Cu^{2+} prefers octahedral or planar coordination in proteins (*95*), tetrahedral coordination is preferred for Zn^{2+} and octahedral coordination is rare (*96*).

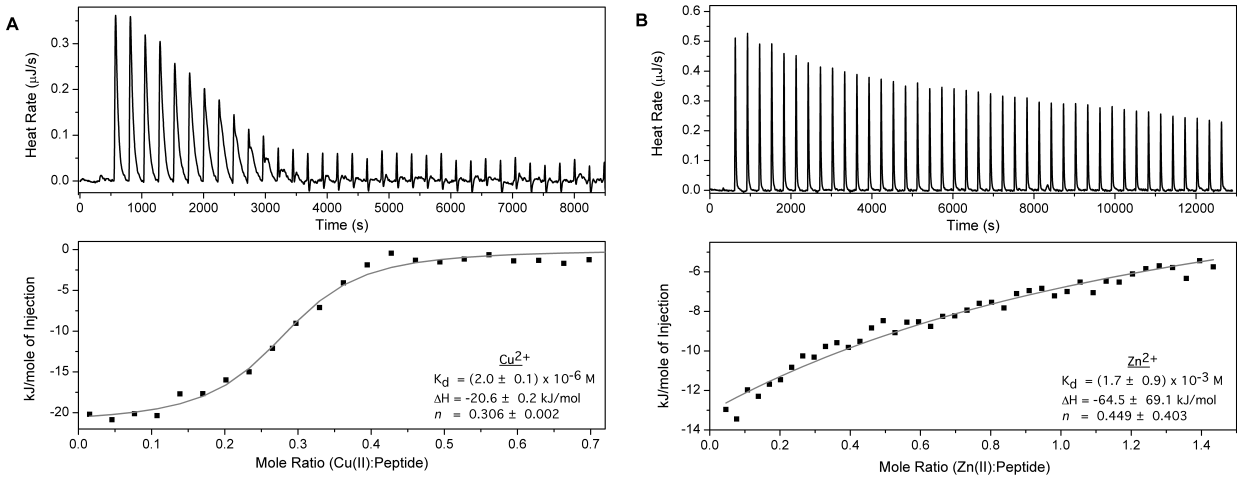

Figure A1-5 ITC experiments. Cu^{2+} (**A**) and Zn^{2+} (**B**) binding to SEVI. Upper panels show the heats evolved during the titrations. Lower panels show the enthalpy changes with increasing metal:SEVI mole fraction. The data were fit to a single binding site (independent binding) model.

Discussion

These data show that Zn^{2+} and Cu^{2+} inhibit SEVI fibril formation (Figures A1-1, A1-2) but do not affect fibril morphology (Figure A1-3). Inhibition is achieved by coordinating the two histidines at positions 3 and 23 in the SEVI sequence (Figure A1-4). Cu^{2+} is more effective in inhibiting SEVI fibrillization than Zn^{2+} and this correlates with the tighter binding of the former metal ion to SEVI (Figure A1-5).

The normal concentrations of Cu^{2+} in human semen is 1-3 μM (*83, 97*), comparable to the K_d of 2 μM determined for binding of the metal to SEVI by ITC (Figure A1-5A). Zinc concentrations in seminal plasma are about 2.5 mM (0.16 g/L) (*83, 85, 97*), higher than in any other body fluids or tissues (*83, 98*). Although Zn^{2+} has a large K_d of 2 mM (Figure A1-5B), the concentration of zinc in semen should be sufficiently high to bind SEVI under physiological conditions. The inhibition of SEVI fibrillization by metals could be relieved when the peptide is transferred to a different environment following intercourse. It has been previously shown that the ability of SEVI to enhance HIV infectivity is retained in the acidic environment of the vaginal tract (*77*). In the vaginal tract, the concentrations of Cu^{2+} and Zn^{2+} are smaller than in semen (*99, 100*) and the acidic pH of 4.2 (*77*) would reduce affinity for these metals due to protonation of the histidine ligands.

Cu^{2+} and/or Zn^{2+} have been reported to inhibit the fibrillization of a number of amyloidogenic peptides and proteins including amylin (*88, 89*), insulin (*92*), Aβ (*101*), and prion protein (*93*). A common theme that emerges from these studies is that the metals inhibit fibrillization by coordinating disordered segments of the polypeptide chains and inducing oligomeric structures that are not competent for fibril assembly. For

amyloidogenic precursors that exist as folded globular proteins, such as transthyretin, a possible strategy to inhibit fibrillization is to develop molecules that bind and thereby stabilize the native state (*102*). The present work, as well as previous work on the inhibition of amyloidogenic proteins by metals (*88, 89, 92, 93*), suggests that compounds that bind and stabilize non-native structures incompatible with fibril assembly may provide an avenue for the development of inhibitors of amyloidogenic polypeptides that are intrinsically unfolded.

Materials and Methods

Preparation of SEVI Peptide

SEVI was purchased as a custom-synthesized peptide from Biopeptide (San Diego, CA). The samples were 95% pure by HPLC and had a peptide content of 70%. Samples were received as lyophilized powders and resuspended in 100% dimethyl sulfoxide (DMSO) to make stock solutions of 100 mg/ml peptide. The stock solutions were kept at -80°C when not in use. For experiments, SEVI stocks were diluted in the desired buffers to a final concentration of 4% (v/v) DMSO. $CuCl_2$, $ZnCl_2$, $MgCl_2$, and $CaCl_2$ were from Sigma. EDTA was from Fisher. All buffers were prepared with Milli-Q water and filtered through a 0.22 μm Millipore filter before use. Buffers for monitoring fibril formation were additionally treated with Chelex-100 to remove trace metals, according to the manufacturer's instructions (Bio-Rad, Hercules CA).

SEVI Fibril Formation Kinetics

Fibril kinetic assays were conducted in white 96-well clear-bottom Corning Costar plates (Lowell, MA) sealed with clear polyester tape to prevent evaporation. Unless otherwise specified, 200 μl samples were used with 2 mg/ml of SEVI in pH 7.4 PBS buffer (8 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 2.7 mM KCl and 140 mM NaCl). The SEVI concentration used for this work is lower than the 5-10 mg/ml typically used in previous studies (*77, 79, 103, 104*). The peptide concentration was chosen as a compromise between peptide cost and the time required to complete fibrillization, since reproducibility worsens with increasing reaction times (*105*). We examined a series of SEVI concentrations between 0.1 and 5 mg/ml and found that the ThT fluorescence

plateaus for the fibrillization reactions scale nearly linearly, except at high peptide concentrations where nonspecific aggregation may start to compete with fibrillization (*92*).

The phosphate buffer used in this work poses the potential problem of precipitating metals, as well as forming complexes which reduce the free metal ion concentration (*106, 107*). The precipitating phosphate (PO_4^{3-}) ion forms with a pK_a of 12.6, so that at a physiological pH of 7.4 its concentration will be only 0.006% of the total phosphate concentration, namely 0.6 μM for 10 mM phosphate buffer (*106*). Consistently, we saw no visible precipitation of the samples even at the highest 1 mM metal concentrations used in our PBS buffer. Because the metals can form soluble complexes with $H_2PO_4^-$ and HPO_4^{2-} ions, however, the effective free metal concentration in solution may be as much as 10-fold lower than the total metal concentration (*107*). As an alternative to phosphate, we conducted experiments in 20 mM Tris with 150 mM NaCl (TBS) buffer at pH 7.4, and saw a dependence on $CuCl_2$ concentration similar to that observed in PBS. The amount of fibrils formed in TBS buffer for control samples that had no metals, however, were about 7-fold lower than in PBS based on the ThT fluorescence plateaus at the end of the reaction. The time for completion of the reactions increased about 6-fold in TBS compared to PBS at the same SEVI peptide concentration. Tris, like phosphate, can associate with divalent metals (*106*). In 20 mM HEPES with 150 mM NaCl pH 7.0, which is a Good buffer (*108*) that should have minimal interactions with metals, we detected a 10-fold lower ThT fluorescence plateau and a 3-fold increase in reaction times. These results are consistent with a recent report that SEVI fibrillization is markedly enhanced in anionic buffers such as phosphate (*86*). We thus

used PBS buffer for SEVI fibrillization studies because it consistently gave the best results, and because this buffer most closely replicates physiological conditions (*85*).

In addition to SEVI peptide and buffer, each fibrillization reaction had a 15 μM concentration of the amyloid-specific dye ThT to detect fibrils (*109*), a 0.02 % (v/v) concentration of NaN_3 to prevent bacterial growth during the fibrillization reactions, and a Teflon bead which improves reproducibility by providing more efficient mixing of the reaction components (*105, 110*). The plates were maintained at a temperature of 37°C throughout the experiments. A continuous agitation speed of 240 rpm was used for the reactions. While agitation is necessary for SEVI to form fibrils (*77, 104*) we found that increasing the agitation speed from 240 to 1,200 rpm had no effects on fibril formation kinetics. Fibrillization reactions were carried out in triplicate to obtain estimates on the uncertainties of kinetic parameters. The fluorescence of ThT was measured every 30 min on a Fluoroskan Ascent plate reader (Franklin, MA) using excitation at 440 nm and emission at 490 nm (*109*). Lag times and rates for fibrillization reactions were obtained from non-linear least-squares fits of the data to a published 6-parameter equation (*111*). Reaction plateaus were determined from the fluorescence maxima when the reactions first reached a steady-state. At longer times, once fibrils have formed, the solution is no longer homogeneous so that the fluorescence values can become unreliable (*112*).

Transmission Electron Microscopy (TEM)

Samples for TEM were prepared from 1 mg/ml solutions of SEVI incubated with continuous agitation for 15 days in PBS buffer at pH 7.4 and a temperature of 37°C. The samples contained 1 mM concentrations of either $ZnCl_2$ or $CuCl_2$. We also looked at

control samples without metals and 5 mM EDTA. ThT fluorescence was used to confirm that all fibrillization reactions had reached the plateau stage prior to imaging by TEM. Samples were applied to 400 mesh carbon coated copper grids and negatively stained using 1% uranyl acetate. The samples were applied in 4 µl aliquots to the grids, rinsed with 1 µl of H_2O, and excess liquid was wicked from the grid using filter paper. To stain the samples, 4 µl of 1% uranyl acetate was applied for 45 sec and the excess stain was wicked off with filter paper. Electron micrographs were obtained using a FEI Tecnai G^2 BioTWIN instrument, which is part of the UConn TEM facility.

NMR Spectroscopy for SEVI

NMR data were collected on a 600 MHz Varian Inova spectrometer equipped with a cryogenic probe. The experiments were preformed on 3 mM samples of SEVI in 90% H_2O/10% D_2O. The pH was adjusted to 5.5 or 7.4 using aliquots of 1 M HCl and NaOH solutions. The temperature for the experiments was 10°C to avoid loss of amide proton signals through solvent exchange. NMR assignments were obtained using 2D experiments from the Varian Protein Pack. Residue-type assignments were obtained from 70 ms mixing time TOCSY experiments and sequential assignments were made from 300 ms mixing time NOESY spectra. These were supplemented with ^{1}H-^{15}N HSQC and Cα-selective ^{1}H-^{13}C HSQC spectra collected on samples at natural isotopic abundance to obtain ^{13}C and ^{15}N assignments. The NMR samples under quiescent conditions were stable at 10°C as monitored by 1D NMR spectra collected before and after 2D NMR experiments. NMR assignments have been deposited in the BioMagResonBank (BMRB) under accession numbers 17924 (SEVI) and 17925 (SEVI with 4 mM $ZnCl_2$). To

characterize line broadening in the presence of $ZnCl_2$ and $CuCl_2$, crosspeak volumes in ^{1}H-^{13}C HSQC and ^{1}H-^{15}N HSQC spectra were measured with the "integrator" utility of the iNMR software (Mestrelab Research).

Isothermal Titration Calorimetry (ITC) for SEVI

Experiments were done at 25°C on a Nano ITC low-volume calorimeter (TA Instruments, New Castle, DE). Samples were prepared in 20 mM Tris buffer at pH 7.5 containing 150 mM NaCl (TBS buffer). For the Cu^{2+} experiment, the SEVI concentration was 200 μM and the starting ligand concentration was 1 mM $CuCl_2$. Due to weaker binding, the experiment with Zn^{2+} necessitated a SEVI concentration of 400 μM and a starting ligand concentration of 2 mM $ZnCl_2$. Since Zn^{2+} has a low solubility at basic pH, the 2 mM $ZnCl_2$ solution in 20 mM TBS buffer was prepared by serial dilution, starting from a 200 mM $ZnCl_2$ stock solution prepared in de-ionized water. All solutions were filtered through 0.2 μm filters and degassed under vacuum for 15 min. Control titration experiments to obtain the heat of dilution were performed by titrating TBS solutions of 1 mM $CuCl_2$ or 2 mM $ZnCl_2$, into TBS buffer without protein. The control experiments showed that contributions from enthalpy changes due to dilution were negligible ($\Delta H < -1.0$ kJ/mol). ITC data were analyzed with NanoAnalyze software (v. 2.1.13).

Reprinted in part from Sheftic et al. (*113*)

Appendix Two. Ionization Constants of Free and Micelle-Bound Amyloid-β

Abstract

Amyloid- beta (Aβ) is thought to contribute to the pathology of Alzheimer's disease and like many amyloid peptides may exert cytotoxic effects at cell membranes. It is therefore important to study the properties of these proteins in membrane-like environments. The membrane mimetics most amenable for high-resolution structural studies are micelle monolayers because at least some of the micelles are sufficiently small for NMR studies. An understanding of the contribution of electrostatic interactions to the structural stability of Aβ in the micelle-bound and intrinsically unfolded native states is critical in achieving a more complete understanding of its structure and dynamics. Owing to the fact that the individual contributions of charged residues as well as individual point charges are prominent forces governing intra- and intermolecular interactions we characterized pK_a values for ionizable residues in free and micelle-bound Aβ using NMR.

Introduction

Electrostatic forces play important roles in protein folding, stability and association (*114*). Charged functional groups can also contribute to enzymatic activity and conformational transitions (*114-117*). A better understanding of electrostatic forces provides information on biological mechanisms ranging from structure formation to molecular recognition, and could lead to improved potential energy functions for protein structure prediction and design (*118-123*).

NMR spectroscopy is unique in that it can measure ionization constants (pK_a values) with site-specific resolution (*124*). The protonated and unprotonated states of ionizable residues are almost always in fast exchange on the NMR timescale (*125*) so that a population-weighted average of the charged and neutral species is observed. At high pH, the population-averaged chemical shift is dominated by the unprotonated form. As the pH is lowered, the resonance position of the titrating site gradually shifts towards the chemical shift of the protonated form, eventually reaching an acidic chemical shift plateau as the population of the protonated form dominates. The pH value at the mid-point of the sigmoidal titration curve is the pK_a value for the site in question. The pK_a value is typically determined from a non-linear least-squares fit of the NMR chemical shift vs pH data to the model specified by the Henderson-Hasselbalch equation (*116, 124*). Although the majority of ionization equilibria are in fast exchange on the NMR timescale, there are some exceptions. Histidine resonances sometimes broaden at pH values near their pK_a consistent with intermediate exchange between the protonated and unprotonated species (*126*). For the vast majority of cases, however, ionization equilibria are in fast exchange on the NMR timescale.

Many of the proteins involved in amyloidogenic diseases such as amyloid-beta (Aβ), amylin, and α-synuclein (αS) function as membrane-bound proteins, or may exert their cytotoxic effects through membrane bound forms (*127, 128*). Characterizing the properties of these proteins in membrane-like environments has become vital. To this end we decided to determine the pK_a values of Aβ-40 (the 40 residue Aβ peptide) in sodium dodecyl sulfate (SDS) micelles. The structure of the peptide bound to SDS micelles is known and other biophysical studies have been performed (*129*). The present study describes the measurement of ionization constants in Aβ by NMR spectroscopy. Inferences are obtained from these experiments into the roles of charges in stabilizing the folded structure of the peptide.

Results

pH Titration of Aβ-40 via NMR

A series of ^{1}H-^{15}N HSQC spectra recorded as a function of pH can provide a straightforward and sensitive way to obtain titration data for charged amino acids (*116, 130*). Figure A2-1 illustrates ^{1}H-^{15}N HSQC spectra recorded during a pH titration of monomeric Aβ-40 free in solution. NMR chemical shifts can be easily measured with very high accuracy relative to an internal reference compound such as DSS or TSP (*66*). From the pH series chemical shifts were derived for ionizable residues in Aβ-40. Figure A2-2 shows representative titration curves for Glu11 in Aβ-40 where chemical shifts were recorded as a function of pH.

Interpretation of pK_a Values in Terms of Stability

The foundation for interpreting pK_a data in terms of protein stability is the thermodynamic linkage cycle shown in Figure A2-3, where F is the folded state, FH is the protonated folded state, U is the unfolded state, and UH is the protonated unfolded state (*131*). The vertical arrows describe the change in free energy on folding or unfolding (ΔG^F or ΔG^U). The horizontal arrows describe equilibria related to the protonation of a residue in the folded (ΔG^{FH}) or unfolded state (ΔG^{UH}). The contribution of a charge to protein stability, is therefore given by equation 1 (*132, 133*):

$$\Delta\Delta G_{titr} = -2.303RT * (pK_a^U - pK_a^N) \quad (1)$$

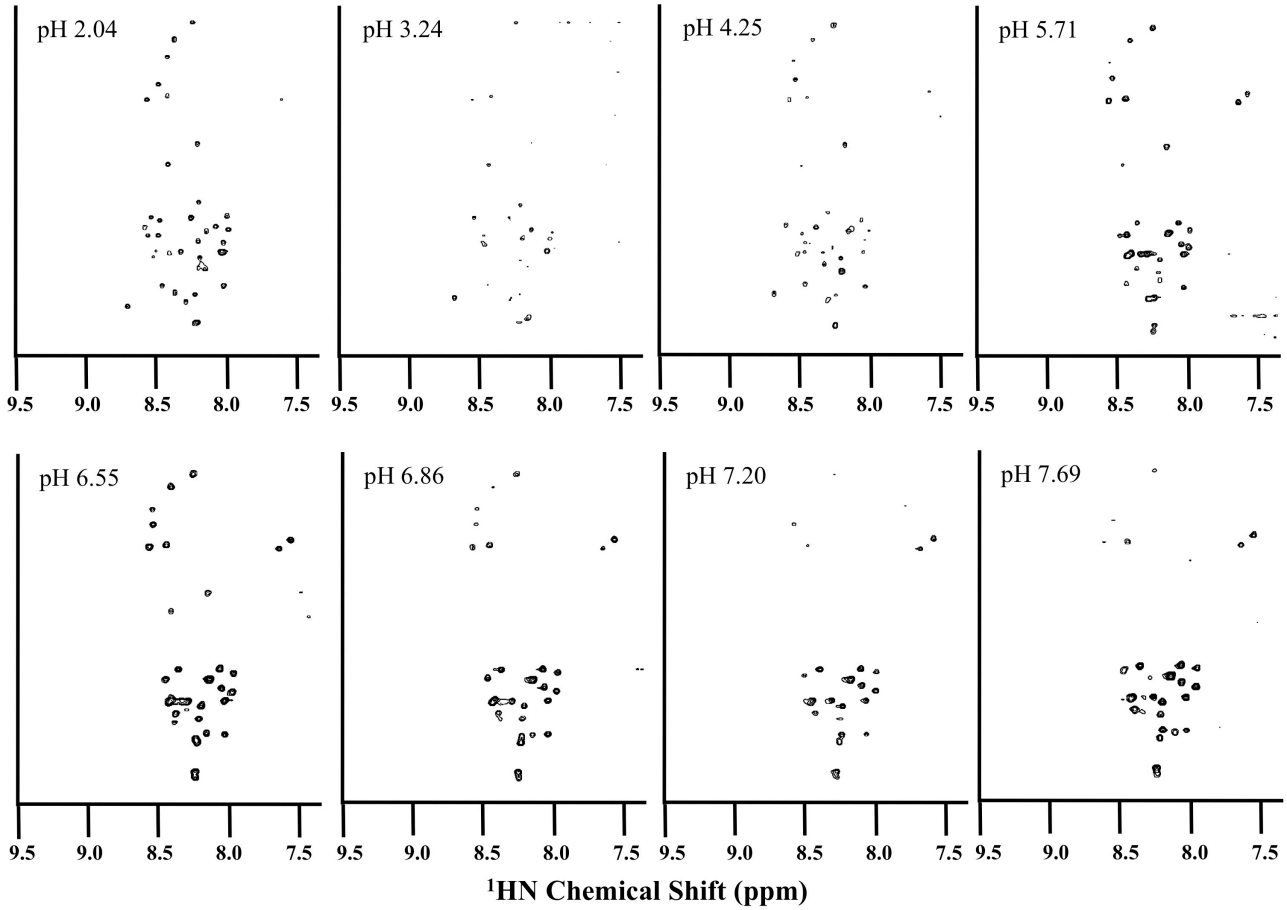

Figure A2-1 ^{1}H-^{15}N HSQC spectra of Aβ-40. Each spectrum was recorded at a different pH and chemical shifts were monitored for perturbations. Note that there are no spectra shown for basic pH values (>8.0) where there is a loss of amide proton peaks because of base-catalyzed exchange with the solvent. All spectra shown were recorded in the absence of SDS micelles at a temperature of 20°C.

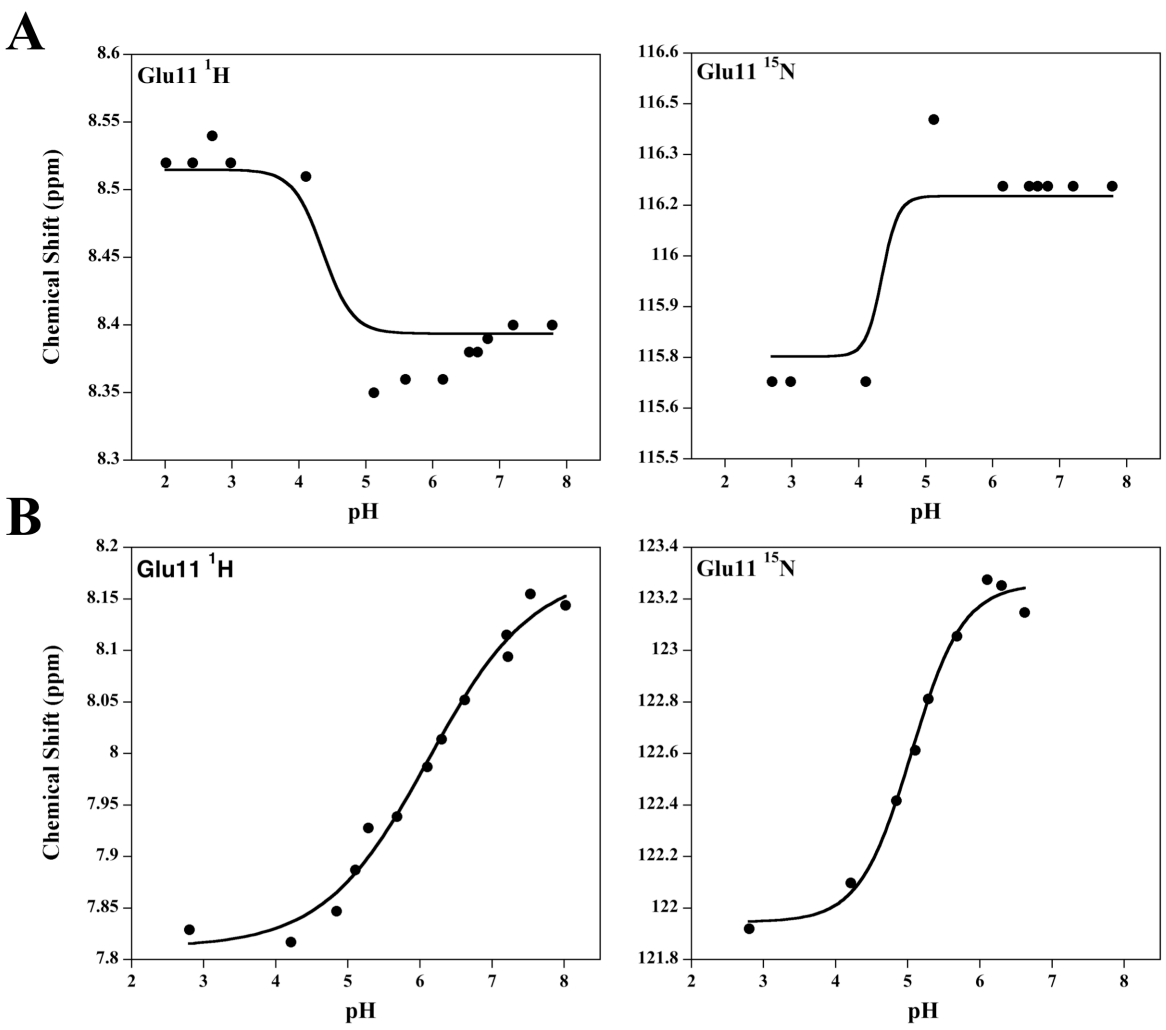

Figure A2-2 Titration of Glu11 using observed chemical shifts as a function of pH. (**A**) Chemical shifts recorded in the absence of SDS micelles for HN and N nuclei. (**B**) Chemical shifts recorded in the presence of 100 mM SDS for HN and N nuclei.

where R is the gas constant, T is absolute temperature, pK_a^U is the pK_a of a site in the unfolded state (Table A2-1) and pK_a^N is the pK_a of a site in the folded state (Table A2-1). The pK_a value of each ionizable site was used to estimate the free energy contribution from protonation to the stability of the folded state using thermodynamic linkage, as shown in Figure A2-3B (*116, 131, 134*).

Results with Micelle Bound Aβ-40

The pK_a data for free Aβ-40 and for the peptide bound to SDS micelles are summarized in Table A2-1. Similar data have been previously obtained for a 1-28 residue fragment of Aβ (*135*). We wanted to see if residues 29-40 in the naturally occurring 40-residue fragment Aβ-40 affected the pK_a values. Our data for Aβ-40 are in very good agreement with those previously reported for the 1-28 fragment. The two exceptions are His6 and Glu11 in the SDS bound form of the peptide. For His6 we think the chemical shift of the high pH form in the previous study (*135*) was probably misassigned to formate, a common and almost ubiquitous impurity in NMR samples that resonates at 8.44 ppm. With our assignment of the high pH plateau of the Hε proton of His6 to 7.7 ppm, the pK_a value for His6 becomes very close to those of the other two histidines. The other small change is that we obtain a pK_a of 6.14 for Glu11 from the pH-dependence of its backbone amide proton chemical shift. In the previous report for the 1-28 fragment a pK_a of 5.5 was obtained from the γ-methylene resonances of Glu11 (*135*). Our correction for Glu11 brings its pK_a closer to those of other glutamates in Aβ-40 (Table A2-1).

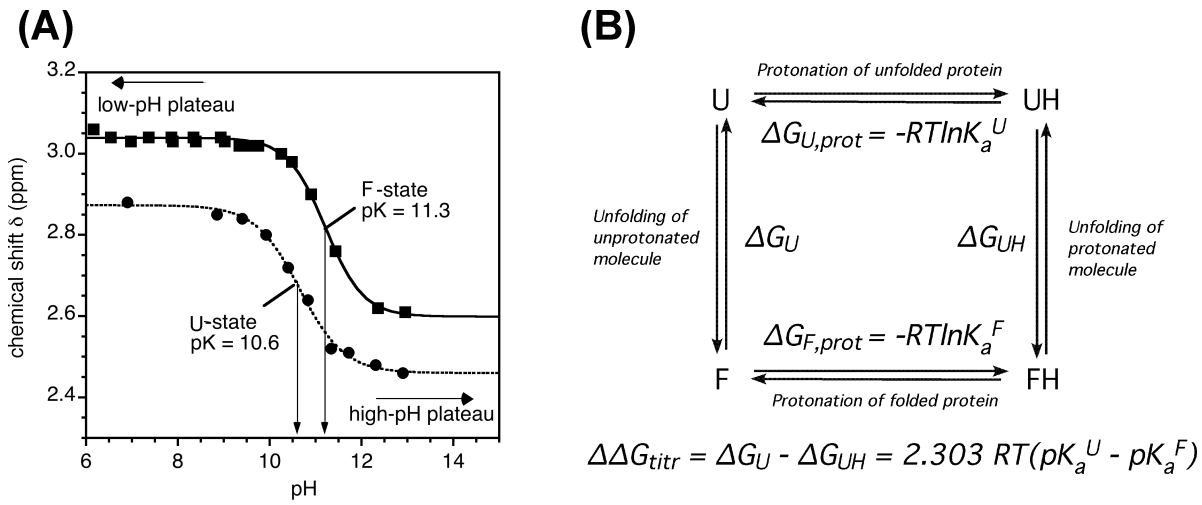

Figure A2-3 Shift in pK_a values between folded and unfolded states and the calculation of a change in folding free energy from a change in charge ($\Delta\Delta G_{titr}$). (**A**) Titration data illustrating the pK_a shift for Lys8 between the folded and unfolded states of the GCN4p coiled coil (*116*). In this case the positively charged state of Lys8 stabilizes the GCN4p structure; the side-chain becomes protonated at a lower hydronium ion concentration in the native state (pK_a 11.3) compared to the unfolded state (pKa 10.6). (**B**) The thermodynamic cycle that can be used to quantify the contribution of a change in charge to a change in stability. U is the unprotonated unfolded state, UH is the protonated unfolded state, F is the unprotonated folded state, FH is the protonated folded state, and R is the gas constant. ΔΔGtitr needs to be multiplied by -1 for acidic groups, and by the number of magnetically equivalent monomers if used for an oligomeric protein. In the present example, for Lys8 in the GCN4p coiled coil dimer ΔΔGtitr = -1.9 (kcal/mol•dimer) at a temperature of 25^{o}C.

The pK_a values for free and micelle bound Aβ-40 as well as the $\Delta\Delta G_{titr}$ values are summarized in Table A2-1. We see large and uniform pK_a shifts when the protein is bound to SDS corresponding to ΔΔG values of a magnitude between 1.2 and 2.7 kcal/mol at 20°C. For groups that become positively charged when ionized, the contributions are favorable likely reflecting attractive interactions with the sulfate headgroups of SDS.

For groups that become negatively charged on ionization the thermodynamic contributions are unfavorable, most likely due to electrostatic repulsion with the negatively charged sulfate headgroups of the micelles. Another system we looked at is the unique histidine at position 50 of αS (Figure A2-4). It is worth noting that αS prefers binding to membranes with a high proportion of negatively charged lipids so the negatively charged SDS micelles in this case should be a better model than neutral DPC micelles (*136*). We identified a large pK_a shift from 6.33 for the free protein to 7.26 for the protein bound to SDS micelles. The larger pK_a in the presence of SDS indicates His50 takes up a proton at smaller hydronium ion concentrations presumably because of electrostatic attraction between the positively charged histidine and the negatively charged sulfate groups on the SDS micelle.

The pK_a shift in His50 of αS corresponds to an energetically favorable $\Delta\Delta G_{titr}$ of -1.3 kcal/mol at 37°C. By contrast in neutral DPC micelles the pK_a is raised only to 6.59, corresponding to a $\Delta\Delta G_{titr}$ of -0.4 kcal/mol. There appears to be a residual interaction, possibly involving a weak ion pair between Glu46 and His50 but its contribution to the stability of the folded structure is much smaller than that between His50 and the negatively charged micelle sulfate groups of SDS. The pK_a values for Aβ-28 bound to DPC micelles have been reported previously (*135*) and for all 10 ionizable residues

investigated large pK_a shifts were observed when the peptide was bound to negatively charged SDS micelles. By contrast, when bound to neutral DPC micelles the pK_a values for Aβ-28 were similar to those of the unstructured peptide in D_2O (*135*).

Table A2-1 Ionization constants for Aβ(1-40) and free energy contributions to the stability of the SDS micelle-bound structure.

Residue	pK_a free	pK_a SDS	$\Delta\Delta G_{titr}$ (kcal/mol)	Resonance	H_2O/D_2O[b]
N-term	8.11 ± 0.64	7.99 ± 0.15	0.16 ± 0.88	Asp 1 ($H\beta_{1,2}$)	D_2O
Asp 1	3.79 ± 0.40	5.43 ± 0.11	2.19 ± 0.56	Ala 2 (H)	H_2O
Glu 3	4.62 ± 0.34	6.17 ± 0.12	2.08 ± 0.48	Glu 3 (H)	H_2O
His 6	6.77 ± 0.03	7.87 ± 0.05	-1.47 ± 0.08	His 6 (Hε1)	D_2O
Asp 7	4.20 ± 0.12	5.55 ± 0.09	1.81 ± 0.20	Asp 7 (H/N)	H_2O
Tyr 10	11.42 ± 0.22	12.30 ± 0.05	1.18 ± 0.30	Tyr 10 ($H\varepsilon_{1,2}$)	D_2O
Glu 11	4.40 ± 0.31	6.15 ± 0.13	2.34 ± 0.45	Glu 11 (H)	H_2O
His 13	6.67 ± 0.02	7.72 ± 0.07	-1.41 ± 0.10	His 13 (Hε1)	D_2O
His 14	6.65 ± 0.02	7.82 ± 0.05	-1.57 ± 0.07	His 14 (Hε1)	D_2O
Lys 16[c]	10.50 ± 0.24	> 12.5	< -2.5	Lys($H\varepsilon_{1,2}$-$H\delta_{1,2}$)	D_2O
Glu 22	4.13 ± 0.07	6.12 ± 0.04	2.66 ± 0.11	Glu 22 (H)	H_2O
Asp 23	4.25 ± 0.10	5.52 ± 0.04	1.70 ± 0.14	Asp 23 (N/H)	H_2O
Lys 28[c]	10.5 ± 0.24	> 12.5	< -2.5	Lys($H\varepsilon_{1,2}$-$H\delta_{1,2}$)	D_2O
C-term	3.72 ± 0.09	5.25 ± 0.02	2.41 ± 0.12	Ala 40 (H/N)	H_2O

All data were collected at a temperature of 20°C in the absence of added salts. For the values of the peptide bound to SDS micelles, the SDS concentration was 100 mM.

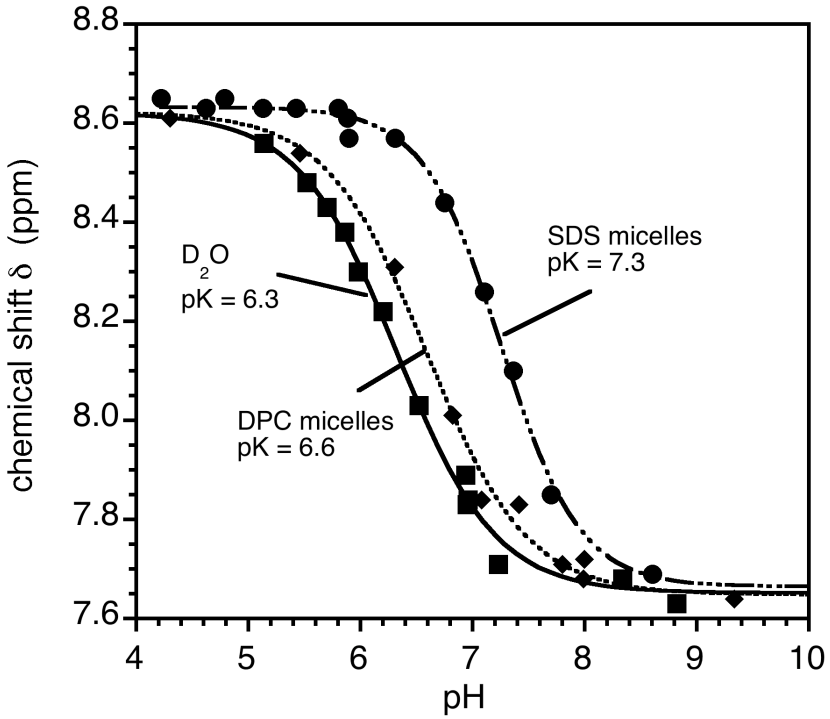

Figure A2-4 Titration curves of His50 from αS in various solvents. All data were collected at 37°C. For the micelle samples the concentrations of SDS and DPC were 100 mM.

Discussion

In summary, binding of proteins to SDS micelles appears to induce large shifts in pK_a values. For negatively charged amino acid sidechains the shifts indicate energetically unfavorable interactions due to electrostatic repulsion. For positively charged amino acid sidechains the pK_a shifts indicate energetically favorable interactions between opposite charges. The magnitudes of the shifts appear to be similar in the different systems we looked at (Aβ and αS), and while they show some dependence on residue type (Table A2-1) they show little dependence on the location of the ionizable group in the amino acid sequence or in the 3-dimensional structure. The pK_a shifts of the glutamate residues investigated are uniformly larger than those of the histidines (Table A2-1). The location of glutamate residues in the sequence, however, did not affect pK_a shifts observed suggesting that the perturbations seen were independent of sequence position. This is important, because residues 1-14 are unstructured in micelle bound Aβ, while residues 15-36 fold into an α-helical structure with a kink between residues 25-27 (*129*). Moreover, paramagnetic relaxation enhancement spin-label experiments show that the segments 15-24 and 29-35 reside on the surface and in the interior of the micelle, respectively; while the unstructured segments 1-14 and 36-40 are solvent exposed. The pK_a shifts for Aβ are insensitive to the differences in structural properties and to the positioning of the different parts of the polypeptide with respect to the micelle (*137*).

Taken together these observations suggest that SDS micelles act as polyanions and provide an electrostatic screen for the entire peptide, that appears to be independent of structure and the degree of immersion of the polypeptides into the micelle. In other words, this screening appears to be non-specific with respect to structural context. In the

case of DPC micelles we expected that if ion pairs within Aβ provided large stabilizing forces we would see large pK_a shifts between the free and bound forms. However, the results with DPC micelles show that pK_a values are similar for bound and free protein (*135*) where the neutral character of these micelles would not provide electrostatic screening. This strongly suggests that electrostatic interactions within the protein are small, compared to those between the protein and the charged micelles.

Materials and Methods

Aβ-40 Preparation for NMR

Natural abundance Aβ-40 was from EZBiolab (Camel, IN) and ^{15}N labeled Aβ-40 was purchased from rPeptide (Bogart, GA). NMR data were collected at a temperature of 20°C on a 600 MHz Varian Inova instrument equipped with a cryogenic probe in the absence of added salts. For experiments with micelles, the SDS concentration was 100 mM. For experiments in H_2O, data were obtained by following the indicated backbone amide resonances using ^{1}H-^{15}N HSQC experiments. At high pH the experiments no longer work because of rapid amide proton exchange. For experiments in D_2O we followed resonances either using 1D ^{1}H NMR for the aromatic residues, or 2D TOCSY for the aliphatic protons. For experiments with micelles we used natural abundance SDS in H_2O and d_{25}-SDS in D_2O.

pK_a Measurements by NMR

To precisely define a pK_a value, a titration series of about 10 – 15 spectra are collected at different pH points. For more demanding experiments it may be possible to obtain a pK_a value with a greater experimental uncertainty from about 6-8 pH points. The most important parts of the titration curve to define are the midpoint, and the acidic and basic plateaus. When planning an experiment it is therefore important to sample pH points near the pK_a and to define the acidic and basic plateaus, by collecting one or two pH points far from the pK_a.

Because of sample losses from pipetting and pH measurements a larger sample volume larger than necessary, 0.7 mL for a standard 5 mm NMR tube, was used. The pH

adjustments were made using stock solutions of acid and base, typically at 1 M concentrations. The acid and base solutions were delivered using 0.2 to 2.0 µl aliquots. For titrations in D_2O, it was necessary to use DCl and NaOD to avoid introducing H_2O into the sample. The pH electrode was also rinsed in 99.8% D_2O before pH measurements, to avoid introducing H_2O. The deuterium isotope effect on the pK_a is about 0.4 pH units. It is known, however, that it is approximately equal and opposite in sign to the deuterium isotope effect on glass electrodes so that no correction was needed for titrations in D_2O compared to H_2O (*124*).

Electrode 2- and 3-point calibrations against pH 4, 7, and 10 buffers were performed for accurate results. The pH of the sample was measured both before and after collecting an NMR spectrum, and the average between these two values was used as the experimental pH. Typically the difference between the pH measured before and after the NMR experiment was within 0.1 to 0.2 pH units. There are two different types of pH electrodes available. Glass electrodes have the advantage of having extensive literature on corrections for the effects of isotopes on the pH measurements and for the effects of denaturants such as urea (*116, 124, 138, 139*). Metal electrodes have the disadvantage that the effects of isotopes or denaturants on pH measurements are not as well characterized as for glass electrodes. Glass electrodes have a particular disadvantage when working with SDS detergent micelles. The potassium salt of SDS is highly insoluble, so even a trace amount of KCl from the storage solution of a glass electrode invariably leads to precipitation of some of the micelles. To avoid any precipitation metal electrodes were used exclusively when studying Aβ-40 in the presence of detergents or lipids.

Analysis of pK_a Values

The data from the pH titration series were analyzed using a non-linear least-squares fit to the Henderson-Hasselbalch equation 2 (*124, 140*):

$$\delta = \delta_{low} - \frac{\delta_{low} - \delta_{high}}{1 + 10^{n(pK_a - pH)}} \quad (2)$$

where pK_a is the ionization constant, δ_{low} is the low pH chemical shift plateau, δ_{high} is the high pH chemical shift plateau, and *n* is the apparent Hill coefficient. With these four parameters it is possible to reconstruct the chemical shift of a titrating nucleus at any pH. The Hill coefficient parameter in principle should describe the cooperativity of the titration (*124*). Values of *n* less than 1 suggest negative cooperativity. Values of *n* greater than 1 suggest positive cooperativity (*124, 141*).

Appendix Three. Visualizing Amyloid-β Complexes

Abstract

The amyloid-beta (Aβ) peptide is linked to the onset and progression of Alzheimer's disease. Two isoforms of the peptide are present in cerebrospinal fluid (Aβ-40 and Aβ-42). The two peptides differ by two C-terminal residues due to variances in sequential proteolysis reactions of the amyloid precursor protein (APP). Despite their strong sequence homology the two peptides have markedly different behavior *in vitro* and *in vivo*. Aβ-42 forms the cross β-sheet amyloid fibril structures at a much faster rate than Aβ-40 whereas Aβ-40 is found in 10-fold higher concentrations *in vivo*. Interestingly, the two peptides also co-localize in amyloid deposits *in vivo*. In order to explore the possible association of the two peptides into a single amyloid fibril total internal reflection fluorescence (TIRF) microscopy was used. The preliminary work here shows that monomeric Aβ-42 forms fibrils that are shorter when grown in the presence of monomeric Aβ-40. It is unclear whether the two peptides co-localize into a single structure but this data shows that Aβ-40 affects the fibrillization of Aβ-42 by possibly acting to inhibit the addition of monomeric Aβ-42 to the growing ends of fibrils once the reaction is in the fibril elongation phase.

Introduction

Proteins can fold into a native conformation as well as misfold into a non-native state. Often proteins can form new misfolded structures that impart detrimental functions. A well-studied class of such proteins is the amyloid family. Amyloids are protein aggregates that have a cross β-sheet structure in which each β-strand is stacked perpendicular to the fibril axis (*142, 143*). Proteinacious aggregates are believed to disrupt normal cellular activity and are often a pathological hallmark of a variety of diseases (*144, 145*) including Parkinson's Disease, Prion disease and Alzheimer's Disease (*146*).

Although it is unclear how these proteins transform into fibrils of ordered β-sheet structure, amyloid fibril formation follows nucleation kinetics (*147, 148*). The kinetics of the reaction can be monitored using ThioflavinT (ThT) assays that can be used to determine variables including reaction rates and lag times (Figure A3-1). ThT is an exogenous fluorescent dye that binds specifically to the cross-beta sheet structure unique to amyloid fibrils (*149*).

The amyloid beta (Aβ) peptide is a byproduct of the sequential proteolytic cleavage of the transmembrane amyloid precursor protein (APP). Two isoforms, the 40-mer and 42-mer, are the most common products of APP cleavage (*150, 151*). Cerebrospinal serum Aβ-42 is detected at about 10-fold lower levels than Aβ-40 but can accelerate the aggregation of Aβ-40 (*152*). It has been shown that altering the ratio of the two peptides by increasing the amount of Aβ-42 relative to Aβ-40 causes a decrease in the lag phase for Aβ-40 fibrillization (*153*). More specifically, when the ratio of Aβ-40: Aβ-42 is 7:3 the toxicity of the fibrils formed is increased (*152*). The amount of Aβ-42

present in cerebrospinal fluid is strongly correlated with Alzheimer's disease. The increase in Aβ-42 relative to Aβ-40 in familial mutants leads to an early onset of the disease (*152*). The ratio of Aβ-42 relative to Aβ-40 is also used as a diagnostic tool for separating Alzheimer's disease from non-Alzheimer's disease related dementia (*154, 155*).

Amyloid fibrils formed by the association of a single protein can have a variety of morphologies, posing the possibility that multiple reaction pathways for aggregation are possible (*156*). It is possible that the two Aβ peptides can be incorporated into a single growing fibril due to their strong sequence similarity. Furthermore, Aβ fibrils formed by the incorporation of both peptides may have different morphologies wherein an increase of Aβ-42 relative to Aβ-40 may prompt the reaction to favor fibrils of a different morphology. The hetero or homogeneity of a fibril during growth can be imaged using total internal reflection fluorescence (TIRF) microscopy (*157, 158*). Determining if both peptides are incorporated into a single fibril may shed light on why changes in the ratio between Aβ-40 and Aβ-42 leads to changes in the rate of fibril formation (*150*). Complimentary studies by TEM can provide information on the morphology of the fibrils formed in the presence of both peptides.

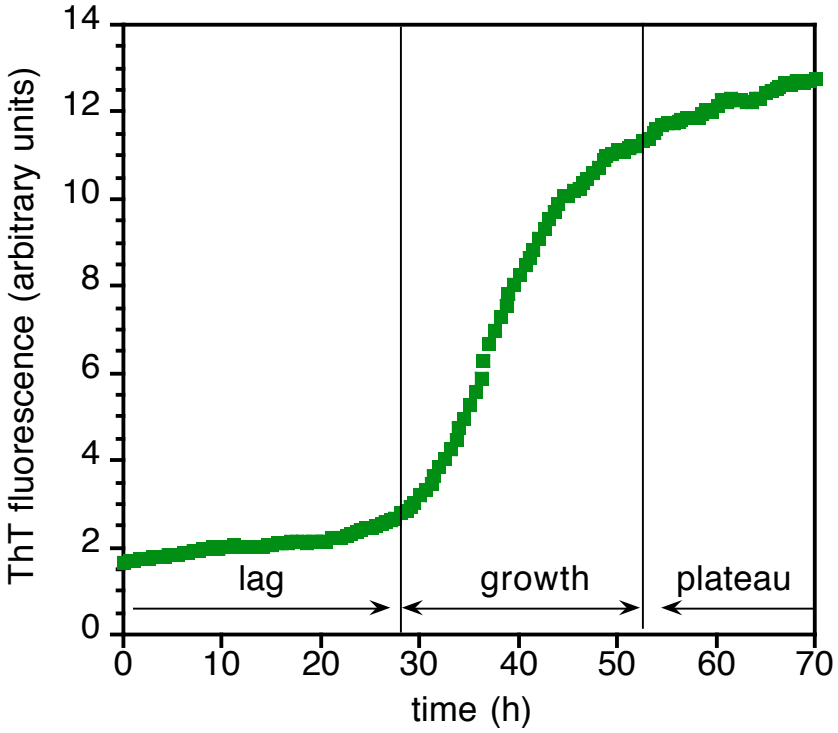

Figure A3-1 Representative amyloid fibrillization reaction monitored by ThT fluorescence. The reaction is characterized by three phases: the lag phase in which the protein remains monomeric until a critical nucleus is formed, the growth phase in which the energetic losses in rotational and translational entropy are compensated by enthalpic increases due to the addition of monomeric protein to a protofibril structure and the plateau phase in which an apparent steady-state between growing fibrils and free monomers is reached.

Results

Microscopy of Aβ-40 and Aβ-42 Peptides

Control experiments were used to generate kinetic reaction curves using ThT assays of the peptides conjugated to fluorescent dyes (Figure A3-2). The Aβ-40 peptide was conjugated to HyLyte Fluor 488 (Aβ-40-488) and Aβ-42 was conjugated to HyLyte Fluor 555 (Aβ-42-555). The sigmoidal transition for Aβ-40-488 (Figure A3-2A) appears inverted when compared to the curve generated from a ThT assay in the absence of other fluorescent dyes (Figure A3-1). The reaction inversion is likely due to Förster resonance energy transfer (FRET) between the ThT and attached HyLyte fluor dye wherein ThT excitation is absorbed by the conjugated 488 dye. A similar inverted profile is seen for Aβ-42-555 (Figure A3-2B) albeit the absence of a detectable lag phase. Due to fast aggregation of Aβ-42 it is likely that the lag time between the preparation of the peptide and the first ThT measurement was longer than the lag time for fibrillization. To confirm the presence of fibrils from reactions using Aβ-40-488 and Aβ-42-555 transmission electron microscopy (TEM) images were collected (Figure A3-3). The conjugated dyes do not appear to affect fibril formation or morphology. For comparison, Aβ-40 fibrils formed without conjugated dyes and in the absence of the exogenous ThT dye are shown in Figure A3-4.

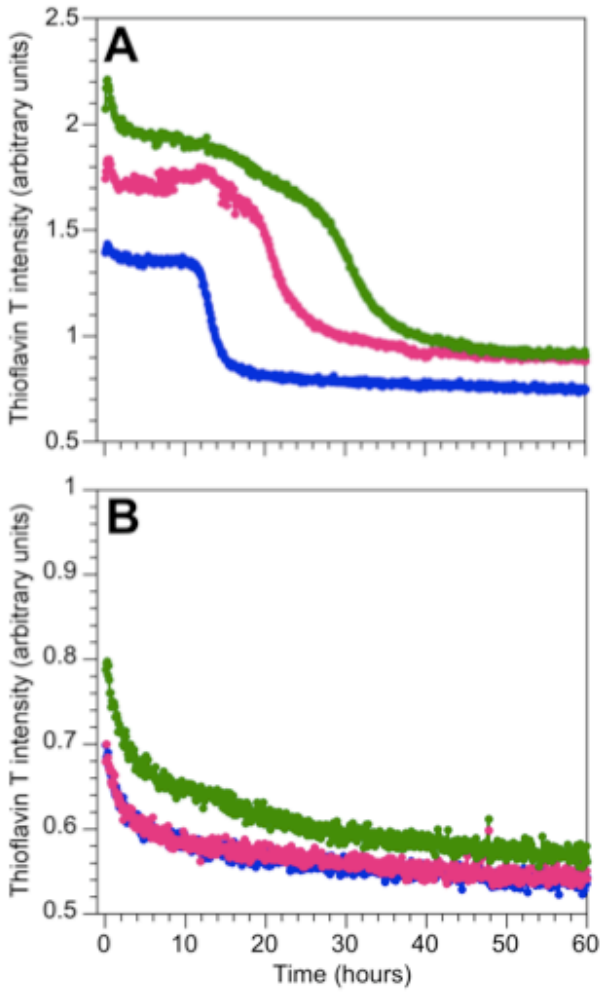

Figure A3-2 ThT kinetic assays for Aβ-40-488 (**A**) and Aβ-42-555 (**B**). Blue curves represent 20 μM and 10 μM peptide concentrations in (**A**) and (**B**) respectively. Pink curves represent 40 μM and 20 μM peptide concentrations in (**A**) and (**B**) respectively. Green curves represent 80 μM and 40 μM peptide concentrations in (**A**) and (**B**) respectively.

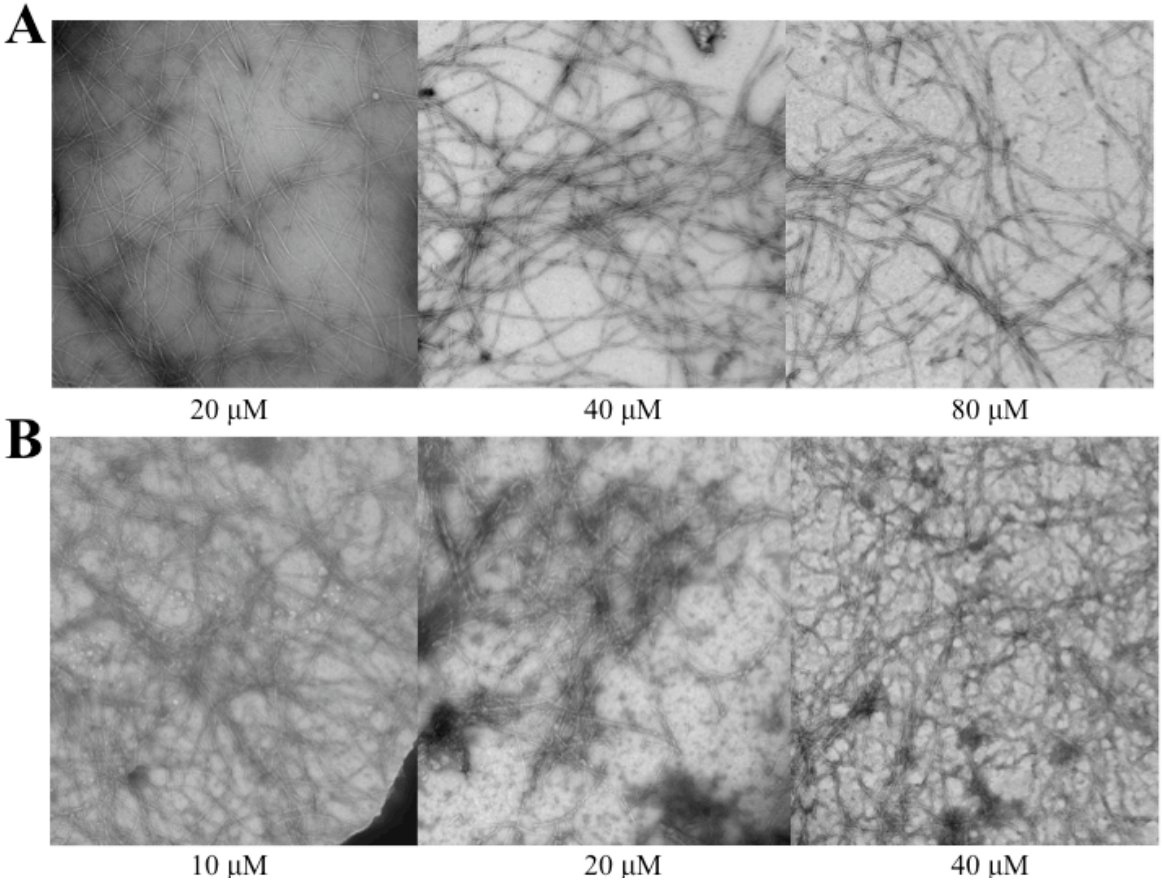

Figure A3-3 TEM images of Aβ-40-488 (**A**) and Aβ-42-555 (**B**). All images were collected at 49,000X magnification. Samples were negatively stained with 1% uranyl acetate on 400 mesh carbon-coated copper grids.

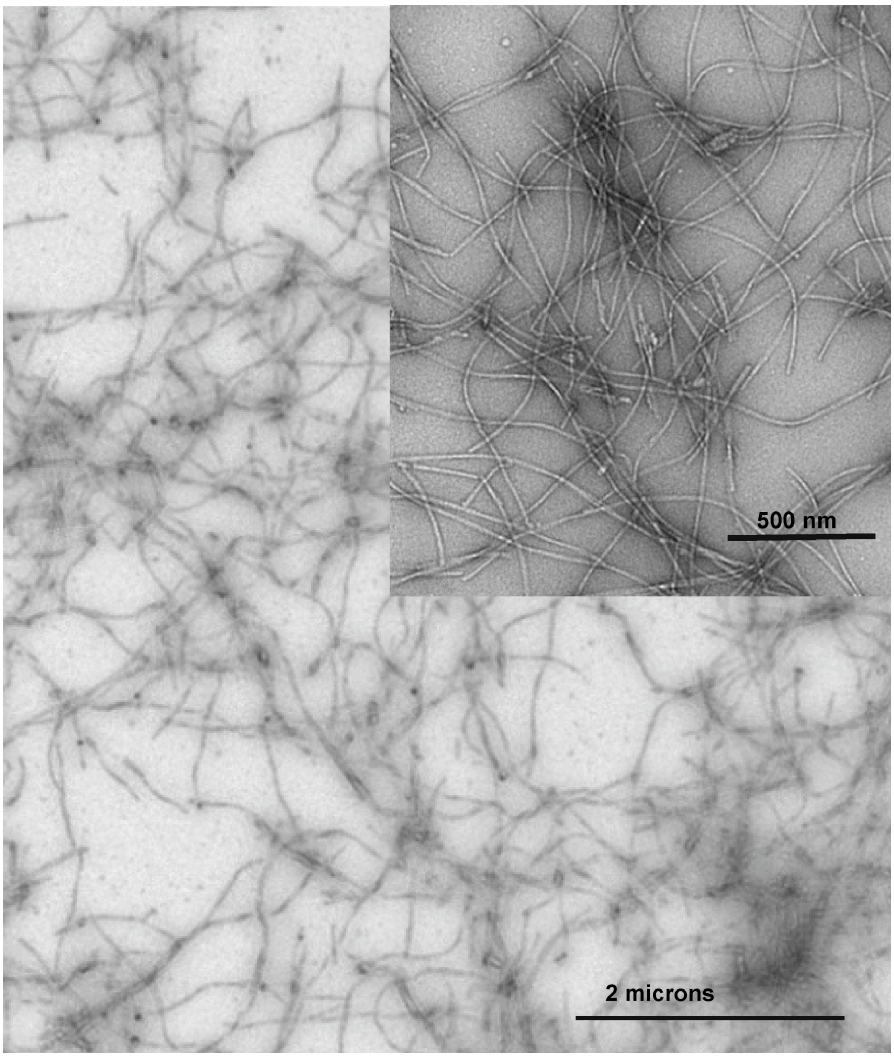

Figure A3-4 TEM images of Aβ-40 fibrils at 11,000X magnification. Fibrils in the inset were imaged at 49,000X magnification. Monomeric Aβ-40 was incubated in 20 mM Na_2HPO_4 pH 7.4 for 14 days prior to imaging. Fibrils were negatively stained with 1% uranyl acetate on 400 mesh carbon-coated copper grids.

TIRF microscopy on Aβ-42-555 shows long extended fibrils similar to those seen in TEM images. TIRF on Aβ-40-488 shows more diffuse fibrils that are difficult to image above the baseline noise (Figure A3-5) and when both peptides are in a reaction mixture it is very difficult to see Aβ-40-488 signal. Due to the diffuse nature of Aβ-40-488 fibrils we used an excess of Aβ-42-555 relative to Aβ-40-488 to obtain better images. The peptide ratio of 2:1 Aβ-42-555:Aβ-40-488 has been successfully imaged using TIRF. In the images in Figure A3-6 it is clear that the length of the Aβ-42-555 fibrils in the 2:1 Aβ-42-555:Aβ-40-488 reaction are shorter when compared to the Aβ-42-555 fibrils prepared alone.

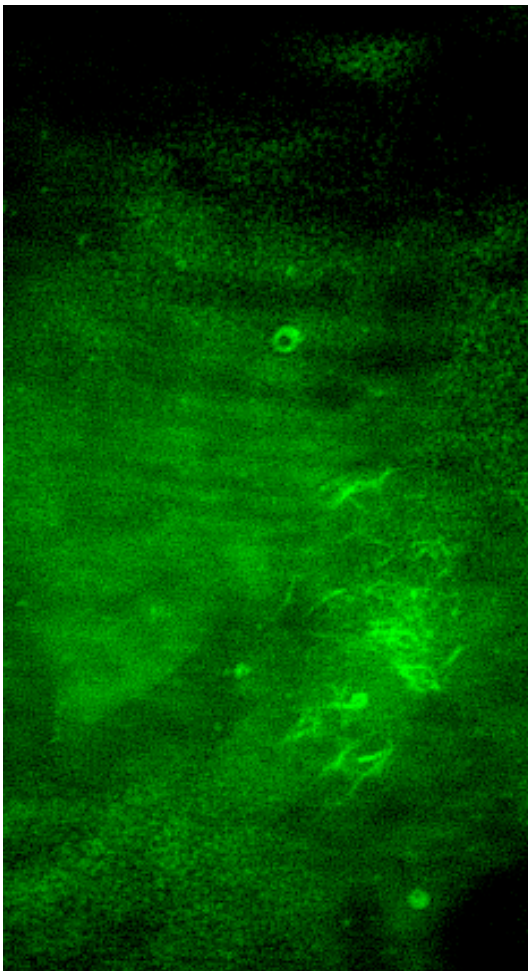

Figure A3-5 TIRF microscopy of Aβ-40-488 fibrils formed in the absence of Aβ-42-555. A monomeric concentration of 40 μM Aβ-40-488 was used to generate the fibrils. Fibrils were formed in a solution containing 20 mM Na_2HPO_4 with 50 mM NaCl and 0.02% NaN_3.

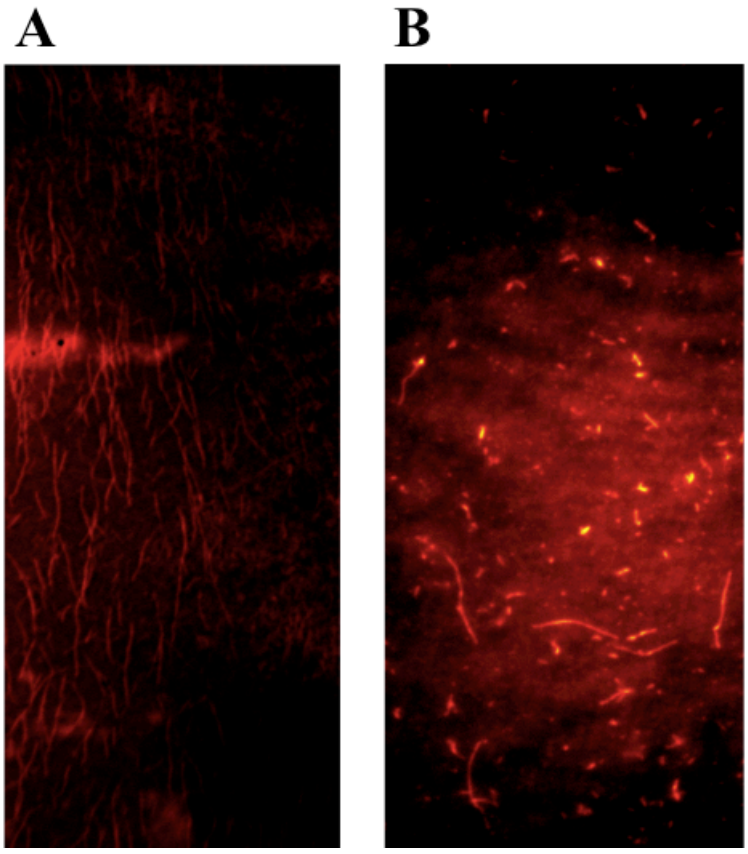

Figure A3-6 TIRF microscopy of Aβ-42-555. (**A**) 20 μM Aβ-42-555. (**B**) 20 μM Aβ-42-555 with 10 μM Aβ-40-488. Both solutions contain 20 mM Na_2HPO_4 with 50 mM NaCl and 0.02% NaN_3.

Discussion

It is unclear at this point whether or not Aβ-40 and Aβ-42 co-localize into a heterogenous fibril. The diffuse fluorescence of Aβ-40-488 by TIRF could be the product of a lesser amount of fibrils formed in the reaction with this isoform. TEM of the Aβ-40-488 peptide suggests that lower quantities of fibrils are formed when compared to the Aβ-42-555 peptide reaction (Figure A3-4). It is likely that the lack of fibrils formed by Aβ-40-488 translates into an inability for the fibril structures present to produce enough fluorescence to overcome background fluorescence in the samples used for TIRF. The enhanced quantity of fibrils formed from Aβ-42-555 visualized by TEM (Figure A3-4) translates into increased signal via TIRF (Figure A3-6).

TIRF of Aβ-42-555 suggests that at least the length of fibrils formed by Aβ-42 are shortened upon addition of Aβ-40. This is consistent with previous work that suggests that the addition of Aβ-40 monomers to Aβ-42 monomers increases the lag time for fibrillization (*159*) and decreases toxicity (*160*). As such the data presented here supports a reaction where nucleation and fibril elongation are affected by the interaction of the two isoforms. The shortened fibrils formed could lessen the density of fibrillar deposits *in vivo*, which could lead to decreased cellular damage particularly at the membrane surface thus decreasing toxicity.

In future studies the diffuse nature of the Aβ-40-488 fibrils should be explored further. The 488 dye itself may be responsible for the diffuse fluorescence seen from TIRF microscopy. One way to test the dye is to switch the 488 and 555 dyes on the Aβ-40 and Aβ-42 peptides, that is to say Aβ-40 should be conjugated to the 555 dye and Aβ-42 should be conjugated to the 488 dye. TIRF microscopy would then reveal if the

diffuse fluorescence is a property of the 488 dye or the Aβ-40 peptide. Furthermore, TEM images of the mixed peptide reactions should be collected to confirm the TIRF finding that Aβ-42 fibril lengths are decreased when they are formed in the presence of monomeric Aβ-40. These studies will lay a foundation for studying the impact of several different ratios of Aβ-40 and Aβ-42 on fibril heterogeneity and morphology.

Materials and Methods

Fluorescently Tagged Aβ Preparation

Samples were prepared from 0.1 mg HyLyte Fluor Aβ-40-488 and 0.1 mg HyLyte Fluor Aβ-42-555 purchased from Anaspec (San Jose, CA). HyLyte Fluor Aβ-40-488 was dissolved in 1% NH_4OH per the manufacturers instructions to a final concentration of 435 uM. Likewise, HyLyte Fluor Aβ-42-555 was dissolved in 1% NH_4OH to a final concentration of 385 uM. The peptide solutions were then sonicated in a cold water bath at 70% amplitude for 5 minutes. Stock solutions were immediately diluted to concentrations of 10-80 μM in white 96-well clear-bottom Corning Costar plates (Lowell, MA) sealed with clear polyester tape to prevent evaporation. Stock solutions were always made fresh on the day of use. Fibrillization reactions were carried out in 20 mM Na_2HPO_4 buffer at pH 7.4 with 50 mM NaCl and 0.02% NaN_3. All solutions were filtered with 25 μm syringe filters. Fibrillization reactions were allowed to progress at room temperature for 20 days before TIRF imaging. For ThT experiments, ThT was added to a final concentration of 15 μM. ThT was purchased from Anaspec.

Aβ Kinetics and TEM

ThT fluorescence assays conducted at 25°C were used to confirm that fibrils were formed from the tagged peptides. The kinetic profiles were inverted from typical reaction curves. Assuming fluorescent interference, namely FRET, was responsible, TEM images were taken of the peptides to confirm fibril formation. Samples were applied to 400 mesh carbon coated copper grids and negatively stained using 1% uranyl acetate. The samples were applied in 4 μl aliquots to the grids, rinsed with 1 μl of H_2O, and excess liquid was

wicked from the grid using filter paper. To stain the samples, 3 µl of 1% uranyl acetate was applied for 45 seconds and the excess stain was wicked off with filter paper. Electron micrographs were obtained using a FEI Tecnai G^2 BioTWIN instrument.

TIRF Chamber Preparation

In order to perform TIRF microscopy, solutions need to be contained in a very thin area or "chamber" created between two different glass surfaces with different refractive indices. In order to create these chambers a glass coverslip (Fisher Cat. No. 12-542-C; Pittsburg, PA) was sealed to a glass slide (Fisher Cat. No. 12-544-I; Pittsburg, PA) leaving a very thing space between the two surfaces. In order to create the chamber space clear sealant was applied to the slide and the coverslip was then placed ontop of the sealant as it dried. The thin space formed between the coverslip and slide was able to hold up to 50 uL of solution. Prior to sealing, the glass surfaces both were cleaned using 1% w/v Alconox detergent and then subsequently rinsed with deionized water. Both surfaces were cleansed with 70% v/v ethanol and allowed to air dry immediately before sealant was applied.

TIRF Microscopy

Samples were prepared as described above albeit no ThT was used for TIRF experiments. Fibril containing samples were injected in 30-40 uL aliquots to chambers created as described above. Slides were inverted for imaging on an Andor Revolution XD Spinning Disk Confocal instrument with a Nikon TiE inverted microscope. Excitation filters of 488 and 560 nm were used to image Aβ-40-488 and Aβ-42-555 respectively.

Images were recorded with an Andor 897 iXon EMCCD camera using an oil immersion TIRF lens (1.49 NA, 100X, Nikon) at a temperature of 25°C.

References

1. Robinson, V. L., Buckler, D. R., and Stock, A. M. (2000) A tale of two components: a novel kinase and a regulatory switch, *Nat Struct Biol 7*, 626-633.
2. Stock, A. M., Robinson, V. L., and Goudreau, P. N. (2000) Two-component signal transduction, *Annu Rev Biochem 69*, 183-215.
3. Stock, J. B., Lukat, G. S., and Stock, A. M. (1991) Bacterial chemotaxis and the molecular logic of intracellular signal transduction networks, *Annu Rev Biophys Biophys Chem 20*, 109-136.
4. Bourret, R. B. (2010) Receiver domain structure and function in response regulator proteins, *Curr Opin Microbiol 13*, 142-149.
5. Jenal, U., and Galperin, M. Y. (2009) Single domain response regulators: molecular switches with emerging roles in cell organization and dynamics, *Curr Opin Microbiol 12*, 152-160.
6. Volkman, B. F., Lipson, D., Wemmer, D. E., and Kern, D. (2001) Two-state allosteric behavior in a single-domain signaling protein, *Science 291*, 2429-2433.
7. Lee, S. Y., Cho, H. S., Pelton, J. G., Yan, D., Berry, E. A., and Wemmer, D. E. (2001) Crystal structure of activated CheY. Comparison with other activated receiver domains, *J Biol Chem 276*, 16425-16431.
8. Stock, A. M., Mottonen, J. M., Stock, J. B., and Schutt, C. E. (1989) Three-dimensional structure of CheY, the response regulator of bacterial chemotaxis, *Nature 337*, 745-749.

9. Foreman, R., Fiebig, A., and Crosson, S. (2012) The LovK-LovR Two-Component System Is a Regulator of the General Stress Pathway in Caulobacter crescentus, *J Bacteriol 194*, 3038-3049.

10. Paul, R., Jaeger, T., Abel, S., Wiederkehr, I., Folcher, M., Biondi, E. G., Laub, M. T., and Jenal, U. (2008) Allosteric regulation of histidine kinases by their cognate response regulator determines cell fate, *Cell 133*, 452-461.

11. Hastings, C. A., Lee, S. Y., Cho, H. S., Yan, D., Kustu, S., and Wemmer, D. E. (2003) High-resolution solution structure of the beryllofluoride-activated NtrC receiver domain, *Biochemistry 42*, 9081-9090.

12. Perez, E., Samper, S., Bordas, Y., Guilhot, C., Gicquel, B., and Martin, C. (2001) An essential role for phoP in Mycobacterium tuberculosis virulence, *Mol Microbiol 41*, 179-187.

13. Cho, H. S., Lee, S. Y., Yan, D., Pan, X., Parkinson, J. S., Kustu, S., Wemmer, D. E., and Pelton, J. G. (2000) NMR structure of activated CheY, *J Mol Biol 297*, 543-551.

14. Kojetin, D. J., Thompson, R. J., Benson, L. M., Naylor, S., Waterman, J., Davies, K. G., Opperman, C. H., Stephenson, K., Hoch, J. A., and Cavanagh, J. (2005) Structural analysis of divalent metals binding to the Bacillus subtilis response regulator Spo0F: the possibility for in vitro metalloregulation in the initiation of sporulation, *Biometals 18*, 449-466.

15. Wright, G. D., Holman, T. R., and Walsh, C. T. (1993) Purification and characterization of VanR and the cytosolic domain of VanS: a two-component

regulatory system required for vancomycin resistance in Enterococcus faecium BM4147, *Biochemistry 32*, 5057-5063.

16. Hess, J. F., Oosawa, K., Kaplan, N., and Simon, M. I. (1988) Phosphorylation of three proteins in the signaling pathway of bacterial chemotaxis, *Cell 53*, 79-87.

17. Garcia, P. P., Bringhurst, R. M., Arango Pinedo, C., and Gage, D. J. (2009) Characterization of a two-component regulatory system that regulates succinate-mediated catabolite repression in Sinorhizobium meliloti, *J Bacteriol 192*, 5725-5735.

18. Karniol, B., and Vierstra, R. D. (2004) The HWE histidine kinases, a new family of bacterial two-component sensor kinases with potentially diverse roles in environmental signaling, *J Bacteriol 186*, 445-453.

19. Feher, V. A., and Cavanagh, J. (1999) Millisecond-timescale motions contribute to the function of the bacterial response regulator protein Spo0F, *Nature 400*, 289-293.

20. Gardino, A. K., and Kern, D. (2007) Functional dynamics of response regulators using NMR relaxation techniques, *Methods Enzymol 423*, 149-165.

21. Stock, J. B., Lukat, G. S., and Stock, A. M. (1991) Bacterial chemotaxis and the molecular logic of intracellular signal transduction networks, *Ann. Rev. Biophys. Biophys. Chem. 20*, 109-136.

22. Stock, A. M., and Guhaniyogi, J. (2006) A new perspective on response regulator activation, *J. Bacteriol. 188*, 7328-7330.

23. Laub, M. T., and Goulian, M. (2007) Specificity in two-component signal transduction pathways, *Ann. Re.v Genet. 41*, 121-145.

24. Stock, A. M., Robinson, V. L., and Goudreau, P. N. (2000) Two-component signal transduction, *Ann. Rev. Biochem. 69*, 183-215.

25. Karniol, B., and Vierstra, R. D. (2004) The HWE histidine kinases, a new family of bacterial two-component sensor kinases with potentially diverse roles in environmental signaling, *J. Bacteriol . 186*, 445-453.

26. Sun, Z. Y., Hyberts, S. G., Rovnyak, D., Park, S., Stern, A. S., Hoch, J. C., and Wagner, G. (2005) High-resolution aliphatic side-chain assignments in 3D HCcoNH experiments with joint H-C evolution and non-uniform sampling, *J Biomol NMR 32*, 55-60.

27. Bryson, K., McGuffin, L. J., Marsden, R. L., Ward, J. J., Sodhi, J. S., and Jones, D. T. (2005) Protein structure prediction servers at University College London, *Nucleic Acids Res 33*, W36-38.

28. Gao, R., and Stock, A. M. (2010) Molecular strategies for phosphorylation-mediated regulation of response regulator activity, *Curr Opin Microbiol 13*, 160-167.

29. Cornilescu, G., Delaglio, F., and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology, *J Biomol NMR 13*, 289-302.

30. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995) 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects, *J Biomol NMR 5*, 67-81.

31. Sheftic, S. R., Garcia, P. P., White, E., Robinson, V. L., Gage, D. J., and Alexandrescu, A. T. (2012) Nuclear magnetic resonance structure and dynamics

of the response regulator Sma0114 from Sinorhizobium meliloti, *Biochemistry 51*, 6932-6941.

32. Dosset, P., Hus, J. C., Blackledge, M., and Marion, D. (2000) Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data, *J Biomol NMR 16*, 23-28.

33. Holm, L., and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D, *Nucleic Acids Res 38*, W545-549.

34. Brunger, A. T. (1996) Xplor Version 3.8 A system for Crystallography and NMR, *Yale University Press, New Haven*.

35. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. and Karplus, M. (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations., *J. Comput. Chem 4*, 187-217.

36. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR, *J Biomol NMR 8*, 477-486.

37. Riepl, H., Scharf, B., Schmitt, R., Kalbitzer, H. R., and Maurer, T. (2004) Solution structures of the inactive and BeF3-activated response regulator CheY2, *J Mol Biol 338*, 287-297.

38. Madhusudan, M., Zapf, J., Hoch, J. A., Whiteley, J. M., Xuong, N. H., and Varughese, K. I. (1997) A response regulatory protein with the site of phosphorylation blocked by an arginine interaction: crystal structure of Spo0F from Bacillus subtilis, *Biochemistry 36*, 12739-12745.

39. Gardino, A. K., Volkman, B. F., Cho, H. S., Lee, S. Y., Wemmer, D. E., and Kern, D. (2003) The NMR solution structure of BeF(3)(-)-activated Spo0F reveals the conformational switch in a phosphorelay system, *J Mol Biol 331*, 245-254.

40. Peters, G. (2009) The effect of Asp54 phosphorylation on the energetics and dynamics in the response regulator protein Spo0F studied by molecular dynamics, *Proteins 75*, 648-658.

41. Kay, L. E., Torchia, D. A., and Bax, A. (1989) Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease, *Biochemistry 28*, 8972-8979.

42. Lipari, G., Szabo, A. (1982) Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 1. Theory and Validity, *J. Am. Chem. Soc. 104*, 4546-4559.

43. Buchan, D. W., Ward, S. M., Lobley, A. E., Nugent, T. C., Bryson, K., and Jones, D. T. (2010) Protein annotation and modelling servers at University College London, *Nucleic Acids Res 38*, W563-568.

44. Stock, A. M., Martinez-Hackert, E., Rasmussen, B. F., West, A. H., Stock, J. B., Ringe, D., and Petsko, G. A. (1993) Structure of the Mg(2+)-bound form of CheY and mechanism of phosphoryl transfer in bacterial chemotaxis, *Biochemistry 32*, 13375-13380.

45. Lewis, R. J., Muchova, K., Brannigan, J. A., Barak, I., Leonard, G., and Wilkinson, A. J. (2000) Domain swapping in the sporulation response regulator Spo0A, *J Mol Biol 297*, 757-770.

46. Menon, S., and Wang, S. (2011) Structure of the response regulator PhoP from Mycobacterium tuberculosis reveals a dimer through the receiver domain, *Biochemistry 50*, 5948-5957.

47. Guhaniyogi, J., Robinson, V. L., and Stock, A. M. (2006) Crystal structures of beryllium fluoride-free and beryllium fluoride-bound CheY in complex with the conserved C-terminal peptide of CheZ reveal dual binding modes specific to CheY conformation, *J Mol Biol 359*, 624-645.

48. Sheftic, S. R., White, E., Gage, D. J., and Alexandrescu, A. T. (2014) NMR Structure of the HWE Kinase Associated Response Regulator Sma0114 in Its Activated State, *Biochemistry 53*, 311-322.

49. Yan, D., Cho, H. S., Hastings, C. A., Igo, M. M., Lee, S. Y., Pelton, J. G., Stewart, V., Wemmer, D. E., and Kustu, S. (1999) Beryllofluoride mimics phosphorylation of NtrC and other bacterial response regulators, *Proc Natl Acad Sci U S A 96*, 14789-14794.

50. Wemmer, D. E., and Kern, D. (2005) Beryllofluoride binding mimics phosphorylation of aspartate in response regulators, *J Bacteriol 187*, 8229-8230.

51. Skerker, J. M., Perchuk, B. S., Siryaporn, A., Lubin, E. A., Ashenberg, O., Goulian, M., and Laub, M. T. (2008) Rewiring the specificity of two-component signal transduction systems, *Cell 133*, 1043-1054.

52. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome, *Proc Natl Acad Sci U S A 98*, 10037-10041.

53. Steinberg, I. Z., and Scheraga, H. A. (1963) Entropy changes accompanying association reactions of proteins, *J Biol Chem 238*, 172-181.

54. Zidek, L., Novotny, M. V., and Stone, M. J. (1999) Increased protein backbone conformational entropy upon hydrophobic ligand binding, *Nat Struct Biol 6*, 1118-1121.

55. Arumugam, S., Gao, G., Patton, B. L., Semenchenko, V., Brew, K., and Van Doren, S. R. (2003) Increased backbone mobility in beta-barrel enhances entropy gain driving binding of N-TIMP-1 to MMP-3, *J Mol Biol 327*, 719-734.

56. Stivers, J. T., Abeygunawardana, C., and Mildvan, A. S. (1996) 15N NMR relaxation studies of free and inhibitor-bound 4-oxalocrotonate tautomerase: backbone dynamics and entropy changes of an enzyme upon inhibitor binding, *Biochemistry 35*, 16036-16047.

57. Alexandrescu, A. T., and Shortle, D. (1994) Backbone dynamics of a highly disordered 131 residue fragment of staphylococcal nuclease, *J Mol Biol 242*, 527-546.

58. Feher, V. A., Zapf, J. W., Hoch, J. A., Whiteley, J. M., McIntosh, L. P., Rance, M., Skelton, N. J., Dahlquist, F. W., and Cavanagh, J. (1997) High-resolution NMR structure and backbone dynamics of the Bacillus subtilis response regulator, Spo0F: implications for phosphorylation and molecular recognition, *Biochemistry 36*, 10015-10025.

59. Stone, M. J. (2001) NMR relaxation studies of the role of conformational entropy in protein stability and ligand binding, *Acc Chem Res 34*, 379-388.

60. Yuan, P., Marshall, V. P., Petzold, G. L., Poorman, R. A., and Stockman, B. J. (1999) Dynamics of stromelysin/inhibitor interactions studied by 15N NMR relaxation measurements: comparison of ligand binding to the S1-S3 and S'1-S'3 subsites, *J Biomol NMR 15*, 55-64.

61. Yu, L., Zhu, C. X., Tse-Dinh, Y. C., and Fesik, S. W. (1996) Backbone dynamics of the C-terminal domain of Escherichia coli topoisomerase I in the absence and presence of single-stranded DNA, *Biochemistry 35*, 9661-9666.

62. Wright, P. E., and Dyson, H. J. (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm, *J Mol Biol 293*, 321-331.

63. Dyson, H. J., and Wright, P. E. (2005) Intrinsically unstructured proteins and their functions, *Nat Rev Mol Cell Biol 6*, 197-208.

64. Cavanagh, J., Fairbrother, Wayne, J., Palmer III, Arthur, G., Rance, Mark., Skelton, Nicholas, J. (2006) *Protein NMR Spectroscopy Principles and Practice (Second Edition)*, Elsevier Inc., Amsterdam.

65. Case, D. A., Dyson, H. J., and Wright, P. E. (1994) Use of chemical shifts and coupling constants in nuclear magnetic resonance structural studies on peptides and proteins, *Methods Enzymol 239*, 392-416.

66. Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., and Sykes, B. D. (1995) 1H, 13C and 15N chemical shift referencing in biomolecular NMR, *J Biomol NMR 6*, 135-140.

67. Vuister, G., Bax, A. (1993) Quantitative J correlation: a new approach for measuring homonuclear three-bond J(HNH.alpha.) coupling constants in 15N-enriched proteins, *Journal of the American Chemical Society 115*, 7772-7777.

68. Ab, E., Schuurman-Wolters, G., Reizer, J., Saier, M. H., Dijkstra, K., Scheek, R. M., and Robillard, G. T. (1997) The NMR side-chain assignments and solution structure of enzyme IIBcellobiose of the phosphoenolpyruvate-dependent phosphotransferase system of Escherichia coli, *Protein Sci 6*, 304-314.

69. Jaravine, V. A., Alexandrescu, A. T., and Grzesiek, S. (2001) Observation of the closing of individual hydrogen bonds during TFE-induced helix formation in a peptide, *Protein Sci 10*, 943-950.

70. Neri, D., Szyperski, T., Otting, G., Senn, H., and Wuthrich, K. (1989) Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional 13C labeling, *Biochemistry 28*, 7510-7516.

71. Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J., and Laue, E. D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline, *Proteins 59*, 687-696.

72. Cates, M. S., Teodoro, M. L., and Phillips, G. N., Jr. (2002) Molecular mechanisms of calcium and magnesium binding to parvalbumin, *Biophys J 82*, 1133-1146.

73. Stites, W. E., Gittis, A. G., Lattman, E. E., and Shortle, D. (1991) In a staphylococcal nuclease mutant the side-chain of a lysine replacing valine 66 is fully buried in the hydrophobic core, *J Mol Biol 221*, 7-14.

74. Nicholson, L. K., Kay, L.E., Torchia, D.A. . (1994) Protein Dynamics as studied by solution NMR techniques. , in *NMR Spectroscopy and its Application to Biomedical Research* (Sarkar, S., Ed.), Elsevier Science Publishers.

75. Ulrich, L. E., and Zhulin, I. B. (2007) MiST: a microbial signal transduction database, *Nucleic Acids Res 35*, D386-390.

76. Sheftic, S. R., Snell, J. M., Jha, S., and Alexandrescu, A. T. (2012) Inhibition of semen-derived enhancer of virus infection (SEVI) fibrillogenesis by zinc and copper, *Eur Biophys J 41*, 695-704.

77. Munch, J., Rucker, E., Standker, L., Adermann, K., Goffinet, C., Schindler, M., Wildum, S., Chinnadurai, R., Rajan, D., Specht, A., Gimenez-Gallego, G., Sanchez, P. C., Fowler, D. M., Koulov, A., Kelly, J. W., Mothes, W., Grivel, J. C., Margolis, L., Keppler, O. T., Forssmann, W. G., and Kirchhoff, F. (2007) Semen-derived amyloid fibrils drastically enhance HIV infection, *Cell 131*, 1059-1071.

78. Kim, K. A., Yolamanova, M., Zirafi, O., Roan, N. R., Staendker, L., Forssmann, W. G., Burgener, A., Dejucq-Rainsford, N., Hahn, B. H., Shaw, G. M., Greene, W. C., Kirchhoff, F., and Munch, J. (2010) Semen-mediated enhancement of HIV infection is donor-dependent and correlates with the levels of SEVI, *Retrovirology 7*, 55.

79. Roan, N. R., Munch, J., Arhel, N., Mothes, W., Neidleman, J., Kobayashi, A., Smith-McCune, K., Kirchhoff, F., and Greene, W. C. (2009) The cationic properties of SEVI underlie its ability to enhance human immunodeficiency virus infection, *J Virol 83*, 73-80.

80. Hong, S., Klein, E. A., Das Gupta, J., Hanke, K., Weight, C. J., Nguyen, C., Gaughan, C., Kim, K. A., Bannert, N., Kirchhoff, F., Munch, J., and Silverman, R. H. (2009) Fibrils of prostatic acid phosphatase fragments boost infections with XMRV (xenotropic murine leukemia virus-related virus), a human retrovirus associated with prostate cancer, *J Virol 83*, 6995-7003.

81. Wurm, M., Schambach, A., Lindemann, D., Hanenberg, H., Standker, L., Forssmann, W. G., Blasczyk, R., and Horn, P. A. (2010) The influence of semen-derived enhancer of virus infection on the efficiency of retroviral gene transfer, *J Gene Med 12*, 137-146.

82. Easterhoff, D., DiMaio, J. T., Doran, T. M., Dewhurst, S., and Nilsson, B. L. (2011) Enhancement of HIV-1 infectivity by simple, self-assembling modular peptides, *Biophys J 100*, 1325-1334.

83. Huang, Y. L., Tseng, W. C., Cheng, S. Y., and Lin, T. H. (2000) Trace elements and lipid peroxidation in human seminal plasma, *Biol Trace Elem Res 76*, 207-215.

84. Martellini, J. A., Cole, A. L., Svoboda, P., Stuchlik, O., Chen, L. M., Chai, K. X., Gangrade, B. K., Sorensen, O. E., Pohl, J., and Cole, A. M. (2011) HIV-1 enhancing effect of prostatic acid phosphatase peptides is reduced in human seminal plasma, *PLoS One 6*, e16285.

85. Owen, D. H., and Katz, D. F. (2005) A review of the physical and chemical properties of human semen and the formulation of a semen simulant, *J Androl 26*, 459-469.

86. Olsen, J. S., Dimaio, J. T., Doran, T. M., Brown, C., Nilsson, B. L., and Dewhurst, S. (2012) Seminal Plasma Accelerates Semen-derived Enhancer of Viral Infection (SEVI) Fibril Formation by the Prostatic Acid Phosphatase (PAP248-286) Peptide, *J Biol Chem 287*, 11842-11849.

87. Tougu, V., Tiiman, A., and Palumaa, P. (2011) Interactions of Zn(II) and Cu(II) ions with Alzheimer's amyloid-beta peptide. Metal ion binding, contribution to fibrillization and toxicity, *Metallomics 3*, 250-261.

88. Brender, J. R., Hartman, K., Nanga, R. P., Popovych, N., de la Salud Bea, R., Vivekanandan, S., Marsh, E. N., and Ramamoorthy, A. (2010) Role of zinc in human islet amyloid polypeptide aggregation, *J Am Chem Soc 132*, 8973-8983.

89. Salamekh, S., Brender, J. R., Hyung, S. J., Nanga, R. P., Vivekanandan, S., Ruotolo, B. T., and Ramamoorthy, A. (2011) A two-site mechanism for the inhibition of IAPP amyloidogenesis by zinc, *J Mol Biol 410*, 294-306.

90. Santner, A., and Uversky, V. N. (2010) Metalloproteomics and metal toxicology of alpha-synuclein, *Metallomics 2*, 378-392.

91. Morgan, C. J., Gelfand, M., Atreya, C., and Miranker, A. D. (2001) Kidney dialysis-associated amyloidosis: a molecular role for copper in fiber formation, *J Mol Biol 309*, 339-345.

92. Noormagi, A., Gavrilova, J., Smirnova, J., Tougu, V., and Palumaa, P. (2010) Zn(II) ions co-secreted with insulin suppress inherent amyloidogenic properties of monomeric insulin, *Biochem J 430*, 511-518.

93. Singh, N., Das, D., Singh, A., and Mohan, M. L. (2010) Prion protein and metal interaction: physiological and pathological implications, *Curr Issues Mol Biol 12*, 99-107.

94. Brender, J. R., Nanga, R. P., Popovych, N., Soong, R., Macdonald, P. M., and Ramamoorthy, A. (2011) The amyloidogenic SEVI precursor, PAP248-286, is highly unfolded in solution despite an underlying helical tendency, *Biochim Biophys Acta 1808*, 1161-1169.

95. Creighton, T. E. (1993) *Proteins: structures and molecular properties*, 2nd ed., W.H. Freeman and Company, New York.

96. Silvennoinen, L., Sandalova, T., and Schneider, G. (2009) The polyketide cyclase RemF from Streptomyces resistomycificus contains an unusual octahedral zinc binding site, *FEBS Lett 583*, 2917-2921.

97. Camejo, M. I., Abdala, L., Vivas-Acevedo, G., Lozano-Hernandez, R., Angeli-Greaves, M., and Greaves, E. D. (2011) Selenium, Copper and Zinc in Seminal Plasma of Men with Varicocele, Relationship with Seminal Parameters, *Biol Trace Elem Res*.

98. Bedwal, R. S., and Bahuguna, A. (1994) Zinc, copper and selenium in reproduction, *Experientia 50*, 626-640.

99. Bohler, K., Meisinger, V., Klade, H., and Reinthaller, A. (1994) Zinc levels of serum and cervicovaginal secretion in recurrent vulvovaginal candidiasis, *Genitourin Med 70*, 308-310.

100. Hagenfeldt, K. (1972) Intrauterine contraception with the copper-T device. 4. Influence on protein and copper concentrations and enzyme activities in uterine washings, *Contraception 6*, 219-230.

101. Raman, B., Ban, T., Yamaguchi, K., Sakai, M., Kawai, T., Naiki, H., and Goto, Y. (2005) Metal ion-dependent effects of clioquinol on the fibril growth of an amyloid {beta} peptide, *J Biol Chem 280*, 16157-16162.

102. Sacchettini, J. C., and Kelly, J. W. (2002) Therapeutic strategies for human amyloid diseases, *Nat Rev Drug Discov 1*, 267-275.

103. Hauber, I., Hohenberg, H., Holstermann, B., Hunstein, W., and Hauber, J. (2009) The main green tea polyphenol epigallocatechin-3-gallate counteracts semen-mediated enhancement of HIV infection, *Proc Natl Acad Sci U S A 106*, 9033-9038.

104. Olsen, J. S., Brown, C., Capule, C. C., Rubinshtein, M., Doran, T. M., Srivastava, R. K., Feng, C., Nilsson, B. L., Yang, J., and Dewhurst, S. (2010) Amyloid-binding small molecules efficiently block SEVI (semen-derived enhancer of virus infection)- and semen-mediated enhancement of HIV-1 infection, *J Biol Chem 285*, 35488-35496.

105. Giehm, L., and Otzen, D. E. (2010) Strategies to increase the reproducibility of protein fibrillization in plate reader assays, *Anal Biochem 400*, 270-281.

106. Aslamkhan, A. G., Aslamkhan, A., and Ahearn, G. A. (2002) Preparation of metal ion buffers for biological experimentation: a methods approach with emphasis on iron and zinc, *J Exp Zool 292*, 507-522.

107. Collier, H. B. (1979) Binding of Zn2+ by buffers, *Clin Chem 25*, 495-496.

108. Good, N. E., Winget, G. D., Winter, W., Connolly, T. N., Izawa, S., and Singh, R. M. (1966) Hydrogen ion buffers for biological research, *Biochemistry 5*, 467-477.

109. LeVine, H., 3rd. (1993) Thioflavine T interaction with synthetic Alzheimer's disease beta-amyloid peptides: detection of amyloid aggregation in solution, *Protein Sci 2*, 404-410.

110. Morris, A. M., Watzky, M. A., and Finke, R. G. (2009) Protein aggregation kinetics, mechanism, and curve-fitting: a review of the literature, *Biochim Biophys Acta 1794*, 375-397.

111. Cohlberg, J. A., Li, J., Uversky, V. N., and Fink, A. L. (2002) Heparin and other glycosaminoglycans stimulate the formation of amyloid fibrils from alpha-synuclein in vitro, *Biochemistry 41*, 1502-1511.

112. Volles, M. J., and Lansbury, P. T., Jr. (2007) Relationships between the sequence of alpha-synuclein and its membrane affinity, fibrillization propensity, and yeast toxicity, *J Mol Biol 366*, 1510-1522.

113. Sheftic, S. R., Croke, R. L., LaRochelle, J. R., and Alexandrescu, A. T. (2009) Electrostatic contributions to the stabilities of native proteins and amyloid complexes, *Methods Enzymol 466*, 233-258.

114. Kumar, S., and Nussinov, R. (2002) Close-range electrostatic interactions in proteins, *Chembiochem 3*, 604-617.

115. Makhatadze, G. I., Loladze, V. V., Ermolenko, D. N., Chen, X., and Thomas, S. T. (2003) Contribution of surface salt bridges to protein stability: guidelines for protein engineering, *J Mol Biol 327*, 1135-1148.

116. Matousek, W. M., Ciani, B., Fitch, C. A., Garcia-Moreno, B., Kammerer, R. A., and Alexandrescu, A. T. (2007) Electrostatic contributions to the stability of the GCN4 leucine zipper structure, *J Mol Biol 374*, 206-219.

117. Graeff, R., Liu, Q., Kriksunov, I. A., Hao, Q., and Lee, H. C. (2006) Acidic residues at the active sites of CD38 and ADP-ribosyl cyclase determine nicotinic acid adenine dinucleotide phosphate (NAADP) synthesis and hydrolysis activities, *J Biol Chem 281*, 28951-28957.

118. Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L., and Baker, D. (2003) Design of a novel globular protein fold with atomic-level accuracy, *Science 302*, 1364-1368.

119. Sternberg, M. J., Hayes, F. R., Russell, A. J., Thomas, P. G., and Fersht, A. R. (1987) Prediction of electrostatic effects of engineering of protein charges, *Nature 330*, 86-88.

120. Sheinerman, F. B., and Honig, B. (2002) On the role of electrostatic interactions in the design of protein-protein interfaces, *J Mol Biol 318*, 161-177.

121. Ripoll, D. R., Vila, J. A., and Scheraga, H. A. (2005) On the orientation of the backbone dipoles in native folds, *Proc Natl Acad Sci U S A 102*, 7559-7564.

122. Jones, S., Shanahan, H. P., Berman, H. M., and Thornton, J. M. (2003) Using electrostatic potentials to predict DNA-binding sites on DNA-binding proteins, *Nucleic Acids Res 31*, 7189-7198.

123. Warshel, A., and Papazyan, A. (1998) Electrostatic effects in macromolecules: fundamental concepts and practical modeling, *Curr Opin Struct Biol 8*, 211-217.

124. Markley, J. L. (1975) Observation of histidine residues in proteins by means of nuclear magnetic resonance spectroscopy, *Accounts of Chemical Research 8*, 70-80.

125. Wuthrich, K. (1986) *NMR of proteins and nucleic acids*, John Wiley & Sons, New York.

126. Alexandrescu, A. T., Broadhurst, R. W., Wormald, C., Chyan, C. L., Baum, J., and Dobson, C. M. (1992) 1H-NMR assignments and local environments of aromatic residues in bovine, human and guinea pig variants of alpha-lactalbumin, *Eur J Biochem 210*, 699-709.

127. Alexandrescu, A. T., and Croke, R. L. (2008) NMR of Amyloidogenic Proteins, in *Protein Misfolding* (O'Doherty, C. B., and Byrne, A. C., Eds.), Nova Science Publishers, Hauppauge NY.

128. Lashuel, H. A., Hartley, D., Petre, B. M., Walz, T., and Lansbury, P. T., Jr. (2002) Neurodegenerative disease: amyloid pores from pathogenic mutations, *Nature 418*, 291.

129. Coles, M., Bicknell, W., Watson, A. A., Fairlie, D. P., and Craik, D. J. (1998) Solution structure of amyloid beta-peptide(1-40) in a water-micelle environment. Is the membrane-spanning domain where we think it is?, *Biochemistry 37*, 11064-11077.

130. Pujato, M., Navarro, A., Versace, R., Mancusso, R., Ghose, R., and Tasayco, M. L. (2006) The pH-dependence of amide chemical shift of Asp/Glu reflects its pKa in intrinsically disordered proteins with only local interactions, *Biochim Biophys Acta 1764*, 1227-1233.

131. Bosshard, H. R., Marti, D. N., and Jelesarov, I. (2004) Protein stabilization by salt bridges: concepts, experimental approaches and clarification of some misunderstandings, *J Mol Recognit 17*, 1-16.

132. Lumb, K. J., and Kim, P. S. (1995) Measurement of interhelical electrostatic interactions in the GCN4 leucine zipper, *Science 268*, 436-439.

133. Anderson, D. E., Becktel, W. J., and Dahlquist, F. W. (1990) pH-induced denaturation of proteins: a single salt bridge contributes 3-5 kcal/mol to the free energy of folding of T4 lysozyme, *Biochemistry 29*, 2403-2408.

134. Lumb, K. J., and Kim, P. S. (1996) Interhelical salt bridges, coiled-coil stability, and specificity of dimerization, *Science 271*, 1136-1138.

135. Ma, K., Cancy, E. L., Zhang, Y., Ray, D. G., Wollenberg, K., and Zagorski, M. G. (1999) Residue-Specific pKa Measurements of the Abeta-Peptide and Mechanism of pH-Induced Amyloid Formation, *J. Am. Chem. Soc. 120*, 8698-8706.

136. Eliezer, D., Kutluay, E., Bussell, R., Jr., and Browne, G. (2001) Conformational properties of alpha-synuclein in its free and lipid-associated states, *J Mol Biol 307*, 1061-1073.

137. Jarvet, J., Danielsson, J., Damberg, P., Oleszczuk, M., and Graslund, A. (2007) Positioning of the Alzheimer Abeta(1-40) peptide in SDS micelles using NMR and paramagnetic probes, *J Biomol NMR 39*, 63-72.

138. Acevedo, O., Guzman-Casado, M., Garcia-Mira, M. M., Ibarra-Molero, B., and Sanchez-Ruiz, J. M. (2002) pH corrections in chemical denaturant solutions, *Anal Biochem 306*, 158-161.

139. Marti, D. N. (2005) Apparent pKa shifts of titratable residues at high denaturant concentration and the impact on protein stability, *Biophys Chem 118*, 88-92.

140. Dames, S. A., Kammerer, R. A., Wiltscheck, R., Engel, J., and Alexandrescu, A. T. (1998) NMR structure of a parallel homotrimeric coiled coil, *Nat Struct Biol 5*, 687-691.

141. Kaslik, G., Westler, W. M., Graf, L., and Markley, J. L. (1999) Properties of the His57-Asp102 dyad of rat trypsin D189S in the zymogen, activated enzyme, and alpha1-proteinase inhibitor complexed forms, *Arch Biochem Biophys 362*, 254-264.

142. Koo, E. H., Lansbury, P. T., Jr., and Kelly, J. W. (1999) Amyloid diseases: abnormal protein aggregation in neurodegeneration, *Proc Natl Acad Sci U S A 96*, 9989-9990.

143. Xing, Y., and Higuchi, K. (2002) Amyloid fibril proteins, *Mech Ageing Dev 123*, 1625-1636.

144. Gregersen, N., Bolund, L., and Bross, P. (2005) Protein misfolding, aggregation, and degradation in disease, *Mol Biotechnol 31*, 141-150.

145. Cuervo, A. M., Wong, E. S., and Martinez-Vicente, M. Protein degradation, aggregation, and misfolding, *Mov Disord 25 Suppl 1*, S49-54.

146. Stefani, M. Protein aggregation diseases: toxicity of soluble prefibrillar aggregates and their clinical significance, *Methods Mol Biol 648*, 25-41.

147. Harper, J. D., and Lansbury, P. T., Jr. (1997) Models of amyloid seeding in Alzheimer's disease and scrapie: mechanistic truths and physiological

consequences of the time-dependent solubility of amyloid proteins, *Annu Rev Biochem 66*, 385-407.

148. Naiki, H., and Gejyo, F. (1999) Kinetic analysis of amyloid fibril formation, *Methods Enzymol 309*, 305-318.

149. Biancalana, M., and Koide, S. Molecular mechanism of Thioflavin-T binding to amyloid fibrils, *Biochim Biophys Acta 1804*, 1405-1412.

150. Jan, A., Gokce, O., Luthi-Carter, R., and Lashuel, H. A. (2008) The ratio of monomeric to aggregated forms of Abeta40 and Abeta42 is an important determinant of amyloid-beta aggregation, fibrillogenesis, and toxicity, *J Biol Chem 283*, 28176-28189.

151. Murray, M. M., Bernstein, S. L., Nyugen, V., Condron, M. M., Teplow, D. B., and Bowers, M. T. (2009) Amyloid beta protein: Abeta40 inhibits Abeta42 oligomerization, *J Am Chem Soc 131*, 6316-6317.

152. Kuperstein, I., Broersen, K., Benilova, I., Rozenski, J., Jonckheere, W., Debulpaep, M., Vandersteen, A., Segers-Nolten, I., Van Der Werf, K., Subramaniam, V., Braeken, D., Callewaert, G., Bartic, C., D'Hooge, R., Martins, I. C., Rousseau, F., Schymkowitz, J., and De Strooper, B. Neurotoxicity of Alzheimer's disease Abeta peptides is induced by small changes in the Abeta42 to Abeta40 ratio, *EMBO J 29*, 3408-3420.

153. Spies, P. E., Slats, D., Sjogren, J. M., Kremer, B. P., Verhey, F. R., Rikkert, M. G., and Verbeek, M. M. The cerebrospinal fluid amyloid beta42/40 ratio in the differentiation of Alzheimer's disease from non-Alzheimer's dementia, *Curr Alzheimer Res 7*, 470-476.

154. Spies, P. E., Slats, D., Sjogren, J. M., Kremer, B. P., Verhey, F. R., Rikkert, M. G., and Verbeek, M. M. (2010) The cerebrospinal fluid amyloid beta42/40 ratio in the differentiation of Alzheimer's disease from non-Alzheimer's dementia, *Curr Alzheimer Res 7*, 470-476.

155. Lewczuk, P., Kornhuber, J., Vanmechelen, E., Peters, O., Heuser, I., Maier, W., Jessen, F., Burger, K., Hampel, H., Frolich, L., Henn, F., Falkai, P., Ruther, E., Jahn, H., Luckhaus, C., Perneczky, R., Schmidtke, K., Schroder, J., Kessler, H., Pantel, J., Gertz, H. J., Vanderstichele, H., de Meyer, G., Shapiro, F., Wolf, S., Bibl, M., and Wiltfang, J. (2010) Amyloid beta peptides in plasma in early diagnosis of Alzheimer's disease: A multicenter study with multiplexing, *Exp Neurol 223*, 366-370.

156. Pedersen, J. S., Andersen, C. B., and Otzen, D. E. Amyloid structure--one but not the same: the many levels of fibrillar polymorphism, *FEBS J 277*, 4591-4601.

157. Ban, T., Yamaguchi, K., and Goto, Y. (2006) Direct observation of amyloid fibril growth, propagation, and adaptation, *Acc Chem Res 39*, 663-670.

158. Ban, T., Hamada, D., Hasegawa, K., Naiki, H., and Goto, Y. (2003) Direct observation of amyloid fibril growth monitored by thioflavin T fluorescence, *J Biol Chem 278*, 16462-16465.

159. Hasegawa, K., Yamaguchi, I., Omata, S., Gejyo, F., and Naiki, H. (1999) Interaction between A beta(1-42) and A beta(1-40) in Alzheimer's beta-amyloid fibril formation in vitro, *Biochemistry 38*, 15514-15521.

160. Yan, Y., and Wang, C. (2007) Abeta40 protects non-toxic Abeta42 monomer from aggregation, *J Mol Biol 369*, 909-916.